DOCUMENT RESUME

ED 040 125 SP 003 859

TITLE Teacher Research on Programmed Instruction. A Collection and Critique of Studies.
INSTITUTION Canadian Teachers' Federation, Ottawa (Ontario). Research Div.
PUB DATE 65
NOTE 54p.
AVAILABLE FROM Canadian Teachers' Federation, 444 MacLaren Street, Ottawa 4, Ontario

EDRS PRICE EDRS Price MF-$0.25 HC-$2.80
DESCRIPTORS Chemistry, *Classroom Research, Effective Teaching, Mathematics, *Programed Instruction, *Program Effectiveness, Research Problems, *Research Skills, *Teachers

ABSTRACT

In order to ascertain the effectiveness of programed instruction as compared to teacher instruction and to evaluate the quality of classroom research by teachers, the Canadian Teachers' Federation offered ten instruction programs on mathematics and chemistry to teachers throughout the country and asked them to evaluate the effectiveness of the programs. A total of 200 teachers ordered programs and 11 wrote complete research reports. Many others wrote brief reports. The major problem confronted by teachers in conducting this research was the limited choice of subjects. It was also found that although most participating teachers were quite well versed in research procedures, some of them tended to draw conclusions not justified by the research design. The overall conclusion of the teachers' reports was that programed instruction is equally as effective as teacher instruction, but not more so. The results from both aspects of this study indicate that research should be done on effective methods of combining programed and teacher instruction, and that more support and assistance should be given to teacher research. (This document includes the 11 research reports written by participating teachers.) (RT)

TEACHER RESEARCH

ON

PROGRAMMED INSTRUCTION

A COLLECTION AND CRITIQUE

OF STUDIES

PREPARED BY THE RESEARCH DIVISION OF THE CANADIAN TEACHERS' FEDERATION

C - 64206

APRIL 1965

ACKNOWLEDGEMENTS

The Canadian Teachers' Federation wishes to acknowledge the assistance of the following organizations and individuals: The Encyclopaedia Britannica Press and the Grolier Society of Canada, for the programs in mathematics which they provided, and Mr. Dennis Ord, of William Lyon Mackenzie Collegiate Institute, Toronto, who provided programs in chemistry. Without this generous assistance, the project in the evaluation of programmed instruction would have been impossible. The fulfilment of the project was also dependent upon the extensive effort and cooperation which many teachers, principals and pupils in all parts of Canada gave to evaluating the programs.

Research Division

CANADIAN TEACHERS' FEDERATION

444 MacLaren Street Ottawa 4, Ontario

December, 1964

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

CHAPTER ONE

INTRODUCTION

Change is perhaps the most significant characteristic of the present-day world. Though change itself is nothing new, the increasing pace of it has made widespread developments in social and economic systems plain for all to see. These changes are accompanied by a substantial expansion of knowledge that creates an accelerating demand for skilled performance and trained intelligence. Given such a social and intellectual climate, it is virtually axiomatic that concomitant pressures for change will be exerted upon schools and, in particular, upon the procedures by which schools turn out their product — the educated student. Even as science and technology have modified patterns of work and leisure in recent years, they now promise to modify the process of instruction.

A notion of long-standing equates instruction with the activity of a teacher — a human mediator between knowledge and the learner. The recent scarcity of qualified teachers, the expansion of knowledge, and the rising demand for education have combined to expand the concept of "teacher" to include any process, material or medium which communicates an idea to a learner. Thus, "X teaches Y to Z" is a general statement about teaching where "Z" is the learner and "Y" the thing he is to learn. The "X" might be a person, a film, an electronic tape, or a book. The problem of instruction thus becomes one of arranging these three elements so that communication is effective and meaningful.

Programmed instruction is one of the new media developed to facilitate learning. The introduction of programmed materials into schools and classrooms thus suggests comparisons between programmed instruction and teacher instruction. Which kind of instruction is more effective? This question, although it is of fundamental importance, may never be answered fully or directly since neither teachers nor programs are single, uniform entities. Not only are there different ways of using programs, even as there are different ways of teaching, but also there are kinds of programs, just as there are kinds of teachers. Unambiguous comparisons of the effectiveness of programmed instruction as opposed to teacher instruction are therefore virtually impossible to achieve.

Believing, however, that critical examination of new techniques and proposals for instruction is essential in education, the Canadian Teachers' Federation arranged to provide several thousand programs to teachers across Canada. During September 1963, brief announcements were placed in teachers' magazines in all provinces asking teachers to make evaluations of programmed instruction. The programs offered included several units excerpted from commercially prepared mathematics programs along with several short teacher-made programs in chemistry. Thus only a very restricted range of topics appropriate to high school grades was covered. Subsequently, some 200 teachers received at their request more than 35,000 programs dealing with ten topics in mathematics and chemistry. In addition to the programs themselves, the teachers were also sent several publications as background for research. One of these examined the place of programmed instruction in education[1] and another outlined procedures suitable for conducting quantitative evaluations of the effectiveness of programmed materials for instruction.[2] In addition, the teachers requesting the programs received a third publication discussing the general rationale and methodology of classroom research on curriculum and instruction.[3]

Thus, the powerful expectation was laid upon the recipients of these programs that they conduct some systematic evaluation of the effectivness of programmed instruction or that they attempt to determine and report upon the problems encountered in using this novel means of instruction. To provide a systematic report of the results obtained by these teachers, an extensive questionnaire was sent to each participant in the project.[4] For most teachers using the programs, a reply to this questionnaire constituted their only report to CTF on the effectiveness of the programs. Several teachers, however, made much more extensive reports of their evaluations. These teachers thus provided concrete evidence that they had on their own initiative conducted, and reported upon, full-scale evaluations which made controlled and objective assessments of programmed instruction. These studies are evidence, therefore, that teachers can conduct research in their classrooms if they are given the resources, the training, the opportunity and the encouragement to carry them out. This publication presents the reports of these teachers as a collection and offers some comment on the design of the studies, the findings, and upon some wider issues in teacher research which are suggested by these projects.

Purposes of the Report

This collection of teacher research on programmed instruction and the comment upon it are intended to serve several purposes. In the first instance, the publication of these reports gives an opportunity to examine teacher research as research. Those who are aware of professional activities in contemporary edu-

---

1 Canadian Teachers' Federation: **Programmed Learning and Its Future in Canada.** 1961. 125 p.
2 Canadian Teachers' Federation: **Evaluating Programmed Instruction Materials Through Action Research.** 1963. 58 p.
3 Canadian Teachers' Federation: **The Role of the Classroom Teacher in Educational Research.** 1961. 75 p.
4 The findings of this survey are to be published shortly under the title **Teachers Evaluate Programmed Instruction.**

cation will know to what extent research in the classroom has been advocated as a means to the improvement of instruction. Considerable effort by those interested in curriculum development has been directed toward the stimulation of teacher-directed research. However, reports of studies which teachers do carry out in their classrooms in response to these interests seldom penetrate far beyond the classroom in which the research was conducted. Consequently, teacher research has frequently been an individual or small group activity and the results and benefits of the research have been limited to the participants themselves. This isolation of teacher research has another important and considerable effect. Research is like any other scholarly or intellectual activity; it is most rewarding and productive when it is cumulative and sequential. Improvement and progress in teacher research are therefore dependent upon improving the communication of research from teacher to teacher. The studies published here are thus presented to provide for them a wider distribution than they would otherwise receive. It is hoped that these studies, unlike most teacher research, will not be read and discussed only by those who knew of the original study and who — as so often is the case — have some concern for, or interest and pride in, the outcomes. Rather, it is the intention that these studies be examined critically by teachers and others who believe that research has a place in the classroom. In this way, research may become a more common activity in the classroom than it now is and the research itself improved in quality, scope and variety.

Another purpose of this publication is to provide a number of examples or models for classroom research. Many teachers wishing to conduct their own research may be unable to find studies which deal with the problems — especially the numerous practical difficulties — which face them in their classrooms. Most research studies are designed by those who have access to considerable resources or who are not limited by the conditions of the classroom. All of the studies reported in this collection were designed for and conducted within conditions which prevail commonly in Canadian schools. The methodology used in these studies is basically simple, direct, and yet flexible. By study the procedure and designs of the research reported here, other teachers may come to recognize the possibilities for research which lie in their own schools and classes. It should not be presumed, of course, that publication of these studies implies that they are to be accepted as models or examples which are without blemish or flaw. Every research study must suffer from a number of inevitable limitations and these presented here are no exception in this regard. Rather they should be accepted as models in the sense that they have successfully come to grips with some of the problems a teacher must face in carrying out classroom research. Other teachers, in reading these studies, will undoubtedly see parallels to their own situation and ways in which the design and procedures of the studies could be improved or strengthened.

A third purpose behind this publication concerns the quality of the research product itself. Teacher research is intended to improve instruction by submitting classroom problems and procedures to scientific analysis. Whether this expectation is in fact fulfilled is an issue on which there is surprisingly little evidence. Behind teacher research there are at least two assumptions — (1) that teachers arrive at different conclusions when they examine an educational problem by research than would be the case if other methods of evaluation were used, and (2) that the decisions about instructional procedures which a teacher reaches as a result of research are sounder or more valid decisions than he would reach without the research. Thus, these reports of research offer some evidence about the kind and quality of teacher thought which research produces. There are no means of discovering what differences this research has made to these teachers and their pupils or whether their opinions about the effectiveness of programmed instruction have been altered by the research. However, these studies do provide insights into the quality of thought which research in the classroom stimulates among teachers who engage in it. For this reason, these studies will be of value to those who are involved with in-service training and professional activity of teachers.

A final purpose implied in this collection involves an examination of the conditions necessary for the organization and conduct of valid research by teachers. The studies reported in this publication vary in length, complexity of design, and presumably in validity as well. Lacking complete data on the conditions under which the studies were completed, it is impossible to examine any relationships between the thoroughness of the research and circumstances in the teacher or his environment which contributed to the adequacy of the research. However, it is possible to use these concrete examples of teacher research as an occasion for examining or speculating about the conditions which ideally would produce the most favourable climate for research in the classroom.

The Programs

Ten programs were offered to teachers for their evaluation. Four of these dealt with topics in mathematics and six with topics in chemistry. Commercial publishers provided the programs in mathematics, while the chemistry programs were the work of a practicing high school teacher who had produced the programs for use in his own classes. A full description of the programs is given in Table 1. The information given in this table is the same as that given to teachers who made inquiries about the availability of programs. From this list, teachers selected programs appropriate to their classes and curriculum.

Table 1. Programs Available for Evaluation by Teachers

Topic	*Publisher*	*Length of Program*	*Grade Level of Material*	*Some Suggested Possible Uses*
Signed Numbers: A traditional approach to the addition, subtraction, multiplication and division of signed numbers.	Temac — Encyclopaedia Britannica Press	500 frames (5-8 hours of student time)	Normally taught in Grade 9	(a) Regular instruction for students viewing work for first time. (b) Remedial work for students who have not grasped the principles involved. (c) Review.
Introduction to Sets: Set terminology, union and intersection, Venn diagrams, Cartesion product.	Temac — Encyclopaedia Britannica Press	237 frames (3-5 hours)	This material appears in Grade 9 in most modern mathematics programs, but it can be taught as early as Grade 7 or as	(a) Regular instruction (as above). (b) Enrichment for bright students in Grades 7, 8 and 9.
Ratio, Proportion and Variation: In this unit, the student learns to compare two quantities by means of a ratio. From ratios, he is led to the setting up and solving of proportions, including problems with similar triangles, shadow problems, and lever problems. The concepts of direct and indirect variation are introduced and various word problems using these concepts are solved.	Grolier Society of Canada	357 frames (3-5 hours)	Material would normally be taught in Grades 9 or 10 algebra	(a) Regular instruction in Grade 9, particularly for non-academic classes. (b) Enrichment in Grades 6, 7 or 8.
Trigonometry: In this unit, the student learns the various relationships involving the sides and angles of right triangles. Given two sides, he is taught how to find the third, using the Pythagorean Theorem. The three fundamental trigonometric ratios are introduced, and the student solves many right triangles using three ratios. He learns how to use trigonometric tables. Finally, he solves many applied problems, including those bringing in the concept of angle of elevation and depression.	Grolier Society of Canada	258 frames (3-5 hours)	Grades 9-11	(a) Regular instruction, particularly a unit in trigonometry for a non-academic stream. (b) Enrichment for Grades 8-9 courses.

Table 1. — Continued

Topic	*Publisher*	*Length of Program*	*Grade Level of Material*	*Some Suggested Possible Uses*
Short Programs for First Course in Chemistry: (1) Chemical Symbols (2) Valencies (3) Formulae Nomenclature of Binary compounds (4) Formulae Nomenclature of Acids and Salts (5) Other Formulae (6) Oxidation and Reduction	Teacher-Constructed: Dennis Ord	Each program has less than 100 frames	Grades 11-12	(a) Regular instruction. (b) Review.

The response to the very modest announcements placed in the teachers' magazines during September, 1963, was substantial. Almost 400 inquiries were received. Some of these were from individuals, but a number were made on behalf of groups of teachers. The actual number of programs distributed is perhaps the best measure of the use which teachers made of the programs. Table 2 contains a summary by province and program title of the programs distributed.

Not all of the inquiries resulted in a subsequent decision to use the programs, as may be seen from the fact that nearly twice as many inquiries were received as orders placed. This apparent lag was due in part to the fact that the actual programs available were devoted to a rather narrow range of topics in the high school curriculum. This selection did not, of course, suit all grades nor all courses of study in various provinces. Moreover, the announcement did not reach teachers until the school year was under way. However, the fact that more than 200 teachers received and used the programs and that an almost equal number of other teachers inquired about them shows a generous and interested response to the issue of programmed instruction and to the request to evaluate it under classroom conditions.

Table 2. Programmed Instruction Inquiries and Orders

			Quantities of Programs Ordered									
Prov.	Inquiries	Orders	*1	2	3	4	5	6	7	8	9	10
B.C.	90	60	1568	1634	1178	1023	935	972	729	830	633	339
Alta.	29	19	788	294	437	235	233	260	270	253	229	208
Sask.	22	13	225	360	257	189	619	618	521	521	479	576
Man.	19	11	180	61	131	130	192	150	128	178	118	150
Ont.	128	63	1125	1178	1367	1351	952	1043	1264	1170	1172	1163
Que.	35	17	432	203	299	130	367	351	389	235	353	302
N.B.	25	11	225	89	181	103	115	138	115	148	115	150
N.S.	14	8	125	61	186	194	132	132	48	48	99	48
P.E.I.	2	1	36	36	—	—	80	80	80	80	80	1
Nfld.	5	2	92	92	92	82	46	12	12	12	12	12
Total	369	205	4796	4008	4128	3437	3671	3756	3556	3475	3290	2949

Total programs sent: 37,066

*1 Signed Numbers
2 Introduction to Sets
3 Ratio, Proportion, and Variation
4 Trigonometry
5 Chemical Symbols
6 Valencies
7 Formulae Nomenclature of Binary Compounds
8 Formulae Nomenclature of Acids and Salts
9 Other Formulae
10 Oxidation and Reduction

The Authors

The authors of the reports presented in Chapter Two of this publication are all classroom teachers or principals. Each is identified by name and school. None of these reports was written specifically for publication. Rather, these summaries were prepared to communicate the results of the program evaluations to a restricted audience which included largely teachers in the school who were close to the experiment and interested in its outcomes. The Canadian Teachers' Federation also received these reports, since it had distributed the programs and expressed a strong interest in evaluations of their effectiveness. The decision to publish the reports arises from the feeling that as a group these reports have a value which none taken singly could have.

In preparing the reports for publication, some editing of the original drafts was required. Complete data and statistical calculations which often accompanied the reports have been omitted. However, the authors' statements of purposes, procedures and conclusions have been retained with only minor amendments of style for greater clarity or to remove reference to local conditions which have no direct relation to the conditions of the research. Such changes as have been made are intended to make the designs of the experiments clearer and to bring the reports into similar formats. In all cases, the flavour of the original writing has been maintained as closely as possible.

All extensive reports received as a result of the CTF project are included in this collection. Many other brief reports were received, but these are not included since they either were not based upon quantitative evaluations or else for various reasons did not include complete descriptions of the evaluations or analyses of their results. It is felt that each of the reports given here can stand on its own as a research study in that each involved conscious and formal arrangements in classrooms to answer a problem through the application of research methodology. Each study stated or implied a problem, involved special arrangements in classrooms for controlled comparisons of achievement, and drew conclusions about the effectiveness of programmed instruction.

For each study, a brief synopsis of the problem, method, and conclusions has been prepared as an introduction to the study. In no case was the synopsis prepared by the original author.

Two of the studies were completed apart from the CTF project in programmed instruction. These, however, are comparable to the other studies, except that they used programmed instruction over longer periods of time and involved somewhat more elaborate procedures than most of the other studies. These two are included with this collection because they are similar to the other studies in topic, purpose and methodology.

The Studies

Eleven studies conducted by teachers on the effectiveness of programmed instruction are presented in the following section. With the publication of this research, teachers may read these studies, appraise them critically, and note both their strengths and weaknesses. Although the reports do not make unanimous assessments of the effectiveness of programmed instruction, their analyses do suggest important insights into the dynamics of instruction. Moreover, the reports provide some guidelines for designing effective research within the limitations of usual school and classroom organization. They offer, as well, an indication of the problems and difficulties a teacher might expect to encounter in conducting classroom research. Study of these reports should therefore be of considerable use to those who wish to learn more about teacher research and to those who would like to increase their own preparedness to do such research by profiting the experience of others. Perhaps the most important feature of the reports, however, may be that they record experiences in classrooms that are seldom open to outside observers. Even teachers, who in the usual course of events do not see beyond their own classrooms, may vicariously participate in the experiences of teachers who have undertaken tasks not usually part of the routine of instruction. The fact that these studies investigate problems posed by a new, promising but unproven medium of instruction makes the research of even greater interest.

CHAPTER TWO

THE RESEARCH REPORTS

CHAPTER TWO

THE RESEARCH REPORTS

A. PROGRAMMED INSTRUCTION IN MATHEMATICS FOR GRADE TEN STUDENTS

by

T. NAKA,

L. V. Rogers Secondary School,

Nelson, B.C.

Two matched groups in Grade 10 mathematics used either programmed or regular instruction to study topics in signed numbers and set theory. In both cases, the group using the programs completed their work more rapidly than the control group. Comparisons in achievement favoured the control group for both programs, although the differences were significantly large only with the instruction on the topic of set theory. It was felt that the most useful applications of programmed instruction could be made in regard to remedial or review work.

Two units of programmed mathematics instruction were obtained for use in an experiment involving two Grade 10 classes. The topics included signed numbers in algebra and an introduction to set theory. The experiment was planned for three reasons:

1. To acquaint teachers in the school with programmed instruction techniques and materials;
2. To observe the effectiveness of programmed instruction;
3. To keep our teachers dynamic and energized through action research.

PURPOSE

The purpose of the experiment was to compare class instruction by programmed learning material and by regular classroom instruction.

PROCEDURES

1. The programmed material involved two units, one in signed numbers and the other in the introduction of set theory. Two Grade 10 classes in mathematics contained 24 students in each class. This number of pupils yielded 14 matched pairs. The criteria used for matching students were (a) age, (b) sex, (c) IQ, and (d) achievement in Grade 9 mathematics. The remaining 20 students (10 students in each class) could not be matched but were included in the statistics.

2. *Signed Numbers*:

   Group A was taught by program.
   Group B was taught by regular classroom teacher.

   Introduction to Set Theory:

   Group A was taught by regular classroom teacher.
   Group B was taught by program.

3. For the unit on signed numbers, tests were given to the groups before and after instruction. For the unit on set theory, a pre-test was not given, as the students had no previous knowledge of this unit.

A null hypothesis was considered, namely, that there is no significant difference in achievement between students taught by programmed material and students taught by a regular classroom teacher. Graphs were prepared showing student achievement in tests given to the following groups:

1. All pupils
2. Complete control class (which included the 14 matched students)
3. Complete experimental class (which included the 14 matched students)
4. Control group involving just the matched pairs
5. Experimental group involving just the matched pairs.

These graphs appear as Figure A-1 and A-2.

OBSERVATIONS

Signed Numbers

1. The experimental group covered the unit in less time than the control group.
2. Achievement went up in both groups as expected. The mean achievement for the control group increased more than for the experimental group. However, this difference was not considered to be significantly large.

3. The instructor of the control group was continuously tempted to cover supplementary work not covered by the program.

4. Students of the control group were made to cover the unit at the pace of the average, whereas the students in the experimental group proceeded individually, thus finishing the unit in varying lengths of time.

Introduction to Set Theory

1. Since both groups had no previous knowledge of set theory, a pre-test was not given.

2. Except for one student, the control group made appreciably greater progress in achievement.

3. The experimental group covered the earlier frames very rapidly but slowed down to a very low pace by the 150th frame. Some students decided it would be wise to start again as they began getting question after question wrong or as they were faced with questions they could not answer.

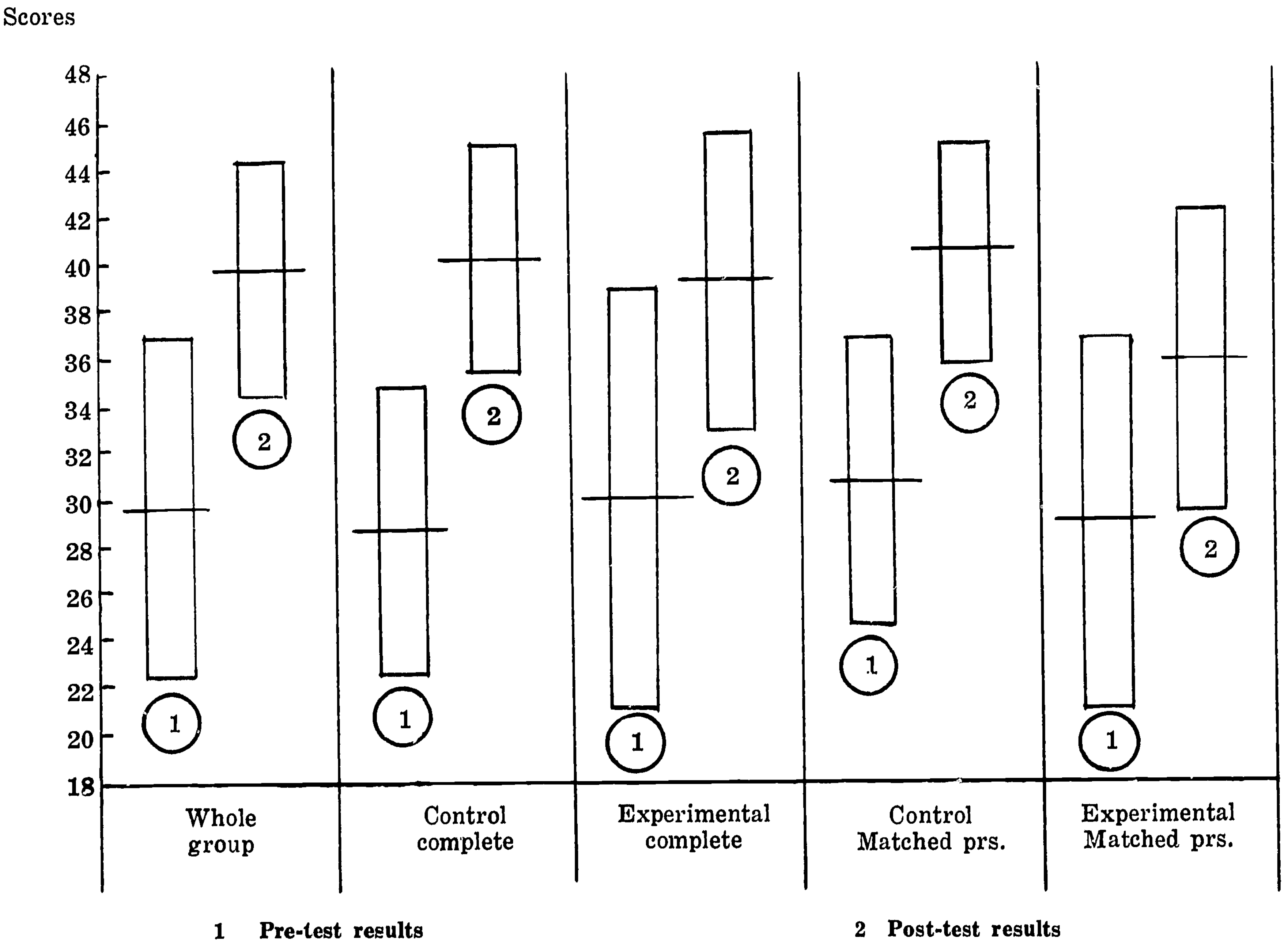

Figure A-1. Comparisons of Groups Using Either Programmed or Regular Instruction for a Unit on Signed Numbers (Pre- and Post-Tests).

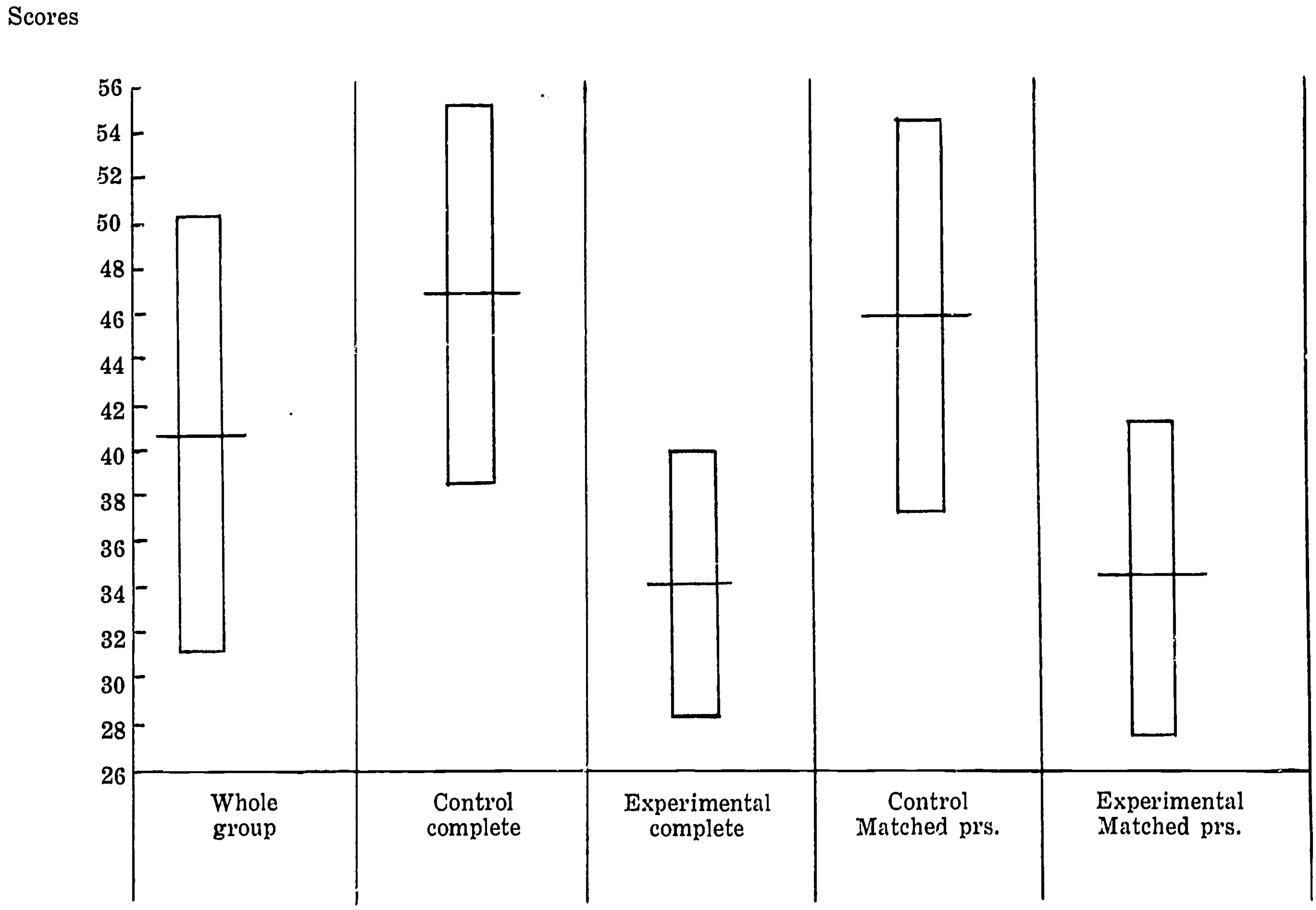

Figure A-2. Comparisons of Groups Using Either Programmed or Regular Instruction for a Unit on Introductory Set Theory (Post-Tests only).

CONCLUSIONS

1. Programmed learning material serves as an excellent review and offers much in remedial instruction. This method does as well as a unit reviewed by a successful classroom teacher.

2. New units or new topics are better taught by a classroom teacher than by programmed material.

It would be wise to point out that our experiment was carried out on a very small scale with low ability students. However, the experiment did give our teachers close contact with programmed learning materials and increased their familiarity with its advantages and limitations. Above all, we have found that programmed instruction can be used with favourable results, if used wisely. We are contemplating further experiments with programmed instruction to endeavour to find its most effective use.

B. AN EXPERIMENT USING PROGRAMMED INSTRUCTION IN MATHEMATICS WITH GRADE SEVEN PUPILS

by

L. C. COOKE and A. C. RUNDLE-WOOLCOCK
Lord Tweedsmuir Elementary School
New Westminster, B.C.

An experimental group of 36 Grade 7 pupils and a similar control group received instruction in elementary algebra. Placement of pupils was made on the basis of intelligence and achievement in arithmetic to create equivalent groups. The experimental groups used only programmed instruction while the control groups received teacher instruction. Measures of instructional effectiveness included comparisons on a composite test and on sub-sections of the test. These sub-tests measured five skills which the instruction aimed to develop. All comparisons favoured the control group over the group using programmed instruction. Methods for improving the effectiveness of programmed instruction grew out of this experimental use of programs.

Grouping

In order to obtain two groups as similar as possible for comparison purposes, the entire Grade 7 group (36 boys and 36 girls) was divided into two groups of 18 boys and 18 girls each.

Placement in groups was made on the basis of mental age as determined by the *Otis Self-administering Tests of Mental Ability, Intermediate Examination: Form C* and computational skill as determined by the *Dominion Survey Test of Arithmetic Fundamentals, Grades 5 to 8, Form A.* Pupils were assigned to groups on the basis of mental age primarily, percentile scores on the arithmetic fundamentals test being the deciding factor in cases of identical or close mental ages. In order to obtain two groups having as nearly identical medians as possible, the groups were matched, but were not randomized, some adjustments being made to preserve as high a statistical equality as possible between the two groups.

As shown in Table B-1, this grouping yielded two groups differing in median mental age by one month. In the percentile scores on the arithmetic test, this grouping yielded indentical medians for the two groups of boys and for the two groups of girls, but gave a difference in medians of the total groups of six percentile points. It was felt that these two groups had as nearly equal potentials and ability as possible (Table B-2).

Table B-1. Mental Ages from Otis Intermediate Examination Form C

	X Group		C Group		X	C	Total Grade
	Boys	Girls	Boys	Girls	Totals		
Median	14.2	15.2	14.3	15.2	14.9	14.8	14.9
Range	16.8-8.5	17.10-11.8	17.2-9.2	17.10-11.0	17.10-8.5	17.10-9.2	

Table B-2. Percentile Ranks from Dominion Survey Test of Arithmetic Fundamentals Form A (Ontario Norms)

	X Group		C Group		X	C	Total Grade
	Boys	Girls	Boys	Girls	Totals		
Median Percentile	65	80	65	80	73.5	67.5	70
Range	98-28	87-10	97-2	99-1	98-10	99-1	

Instructional Procedure

In order to evaluate the program as to efficiency rather than as to speed of learning, both groups were given 11 fifty-minute periods including one period for final testing. Because the grouping involved platooning half classes and because the control group was given no home assignments, the time given both groups was equal.

The program used was "Ratio, Proportion, and Variation" from *Algebra II* of TMI-Grolier. The experimental group proceeded with the program, receiving no other instruction from the teacher than that required by individuals as they encountered difficulty.

The control group received total group instruction on the content of the unit as outlined in the B.C. program of studies plus sufficient content to cover the course as presented in the program. It must be noted here that the program contained more than adequate content for the Grade 7 program in this province. Thus, both groups were presented the same material, the essential difference being the manner of presentation.

Evaluation

Both groups were administered the same five-part teacher-made test, which measured each of the skills involved in the unit separately. The five sections of the test measured the following skills and concepts:

1. Expressing ratios
2. Concept of supplying multipliers to cancel
3. Solving proportions
4. Sums of ratios
5. Setting out and solving proportions.

In the section on solving ratio, marks were awarded for method used as well as for the correct answer. As shown in Table B-3, the median for the total "C" group was 21 per cent higher than for the total "X" group.

Table B-3. Total Test — Including Marks for Method Used in Problem-Solving

	X Group		C Group		X	C	Total
	Boys	Girls	Boys	Girls	Totals		Grade
Median	39%	56%	63%	70%	43%	64%	56%
Range	84%-4%	96%-18%	88%-22%	98%-22%	96%-4%	98%-22%	

In order to ascertain if this marked difference in achievement was due to the possibility that methodology received more stress in the "C" group than in the "X" group, the test scores were also determined on the basis of answers only, no marks being awarded for method. As shown in Table B-4, the marks for the control group were again superior in every category, although the difference in median scores for the two groups was lessened to 13 per cent.

Table B-4. Total Test — Excluding Marks for Method Used in Problem-Solving

	X Group		C Group		X	C	Total
	Boys	Girls	Boys	Girls	Totals		Grade
Median	45%	63%	62%	74%	53%	66%	62%
Range	84%-5%	95%-24%	89%-26%	97%-29%	95%-5%	97%-26%	

To analyze the results further, similar comparisons were made for each of the sub-tests (see Tables B-5 to B-10). These results are given here for the total groups only, although data for boys and girls separately were also analyzed.

In Section 1 (see Table B-5), the median for the total "C" group was 10 per cent higher than that for the "X" group.

In Section 2 (see Table B-6), the largest difference in medians occurred, "C" group's median being 50 per cent higher than that of "X" group, and the results of "X" group in no case exceeding those of "C" group.

Table B-5. Section 1 — Expressive Ratios

	Total Groups	
	X	C
Median	70%	80%
Range	100%-0%	100%-0%

Table B-6. Section 2 — Concept of Supplying Multipliers to Cancel

	Total Groups	
	X	C
Median	30%	80%
Range	100%-0%	100%-40%

Section 3 (solving proportions), contained questions involving the essential computation for which this unit is taught. The only case in which "X" group results were better than those of "C" group was in the lowest score (see Table B-7).

In Section 4 (see Table B-8), the results obtained by the experimental group on programmed learning were equal to those obtained by "C" group students on conventional instruction.

Table B-7. Section 3 — Solving Proportions

	Total Groups	
	X	C
Median	50%	70%
Range	90%-10%	100%-0%

Table B-8. Sums of Ratios

	Total Groups	
	X	C
Median	67%	67%
Range	100%-0%	100%-0%

Section 5, problem-solving by ratio and proportion, contained the most difficult questions and was the most important section of the test. Again, scores were compared both with and without marks assigned for method. As shown in Tables B-9 and B-10, the control group's median was 47 per cent higher when method marks were included and 28 per cent higher

Table B-9. Section 5 — Setting and Solving Proportions Marks for Answers and Method

	Total Groups	
	X	C
Median	16%	63%
Range	100%-0%	100%-11%

Table B-10. Section 5 — Setting and Solving Proportions — Answers Only

	Total Groups	
	X	C
Median	29%	57%
Range	100%-0%	100%-29%

Figures B-1 and B-2 show the relationship of mental age to achievement for both groups, including and excluding marks for method respectively. When marks were awarded for method, the average scores for the control group were much superior for all mental age quartiles. It should be noted, however, that the average scores for the second and third quartile groups on programmed learning were nearly equivalent to each other.

When marks were awarded for answer only (see Figure B-2), the averages of the four quartile groups on traditional instruction were still superior to those of similar groups on programmed learning. In the second quartile, however, the averages of the programmed learning group were nearly equivalent to those of the group on traditional instruction. It should be noted, too, that the averages of the second and third quartile groups are equivalent to those of the fourth quartile group. It would seem, on the basis of this comparison, that programmed instruction was of most benefit to the pupils in the second quartile, or those slightly below average in ability.

Conclusions

1. The results obtained in this experiment would indicate that the achievement of pupils taught by traditional methods is much superior to that obtained by pupils taught by programmed instruction.

2. The program did not provide the students with satisfactory instruction in algebraic methodology, which should be a major outcome of this unit when taught at the Grade 7 level.

3. Students on this learning program were prone to read back through a question incorrectly answered so as to fit the answer to it rather than to re-work the question to obtain the correct answer.

Recommendations

1. That in future, when conducting such experiments, pupils be allowed to become familiar with programmed instruction by means of another unit not to be compared with traditional instruction. The added familiarity with programmed instruction when commencing the experimental program would eliminate differences in achievement that are caused, we feel, more by lack of familiarity with programmed instruction itself than with a fault in the program.

2. That to ensure each pupil's satisfactory progress, periodic test questions to which only the teacher has the answers be included in the program. These could be spaced at increasing intervals as they progress through the program.

3. That the program include more drill on the total concept after all developmental processes have been mastered.

4. That the program concentrate more on developing methodology in problem-solving.

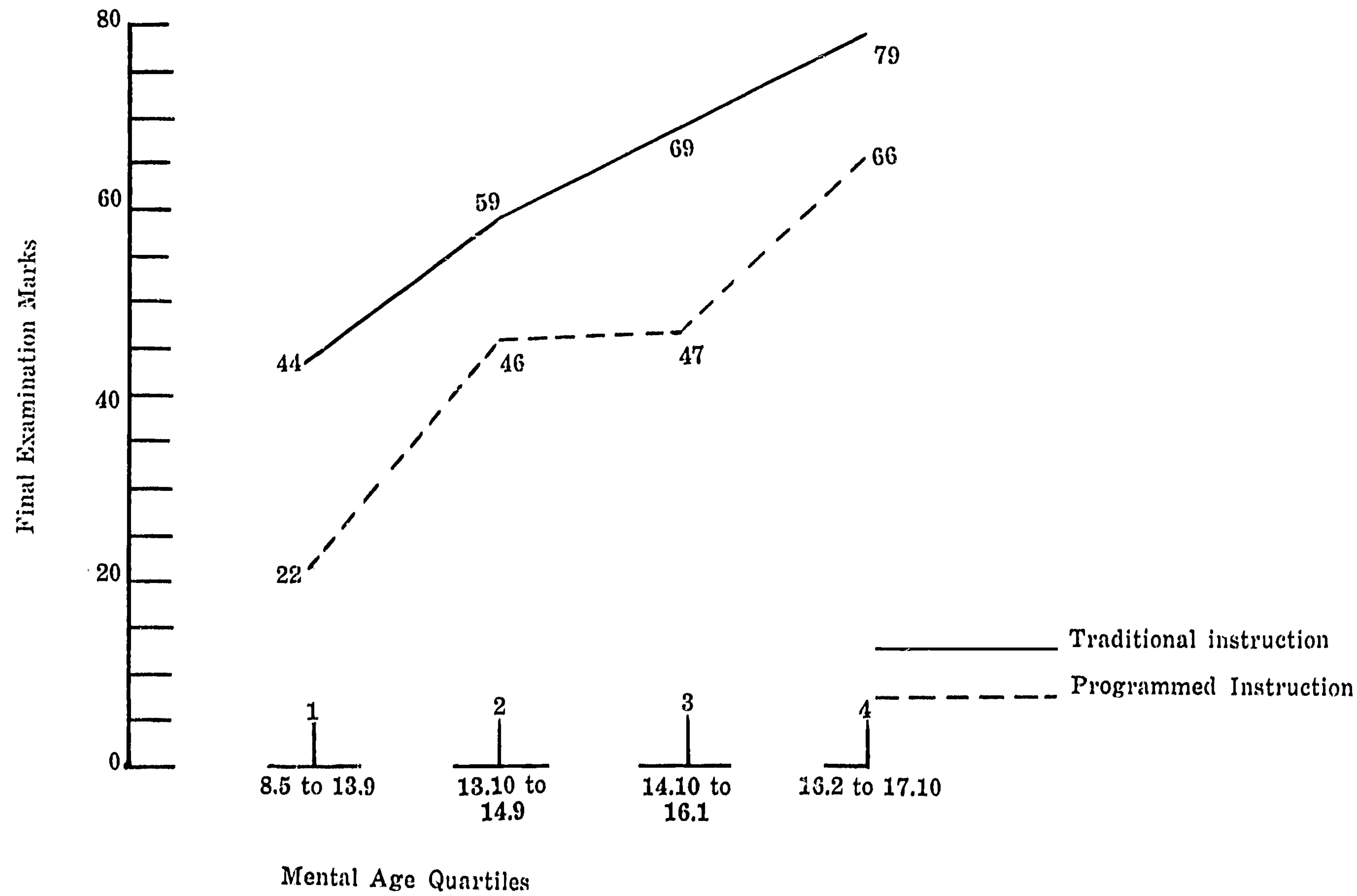

Figure B-1. Relationship of Mental Age to Achievement for Programmed Instruction and Traditional Instruction (scores include marks for method).

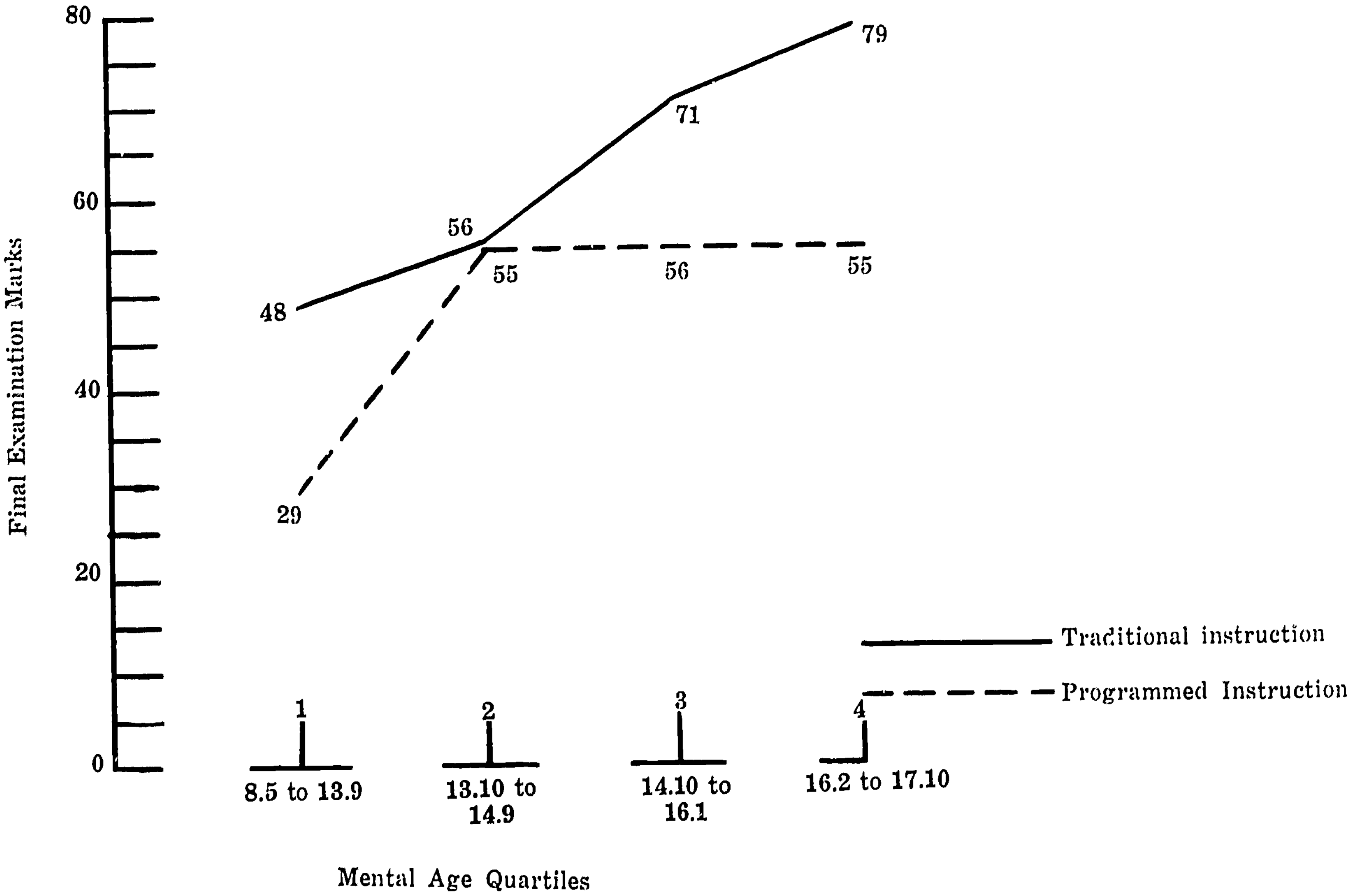

Figure B-2. Relationship of Mental Age to Achievement for Programmed Instruction and Traditional Instruction (scores awarded for answer only).

C. USING PROGRAMMED INSTRUCTION WITH VOCATIONAL STUDENTS OF LOWER ABILITY

by

F. FRANCIS
Victoria Secondary School
Victoria, B.C.

Thirty-four pupils in Grade 10 and twenty-nine in Grade 11 received programmed instruction in mathematics. Both classes were composed of boys of lower ability enrolled in a vocational course. Comparison classes for both grades were formed from boys in the same course and of similar ability. The control classes studied the same topics covered in the programs but received regular classroom instruction. In Grade 10, the achievement of pupils using the programmed materials was lower than pupils in the control classes. In Grade 11, achievement with the programmed materials fell below expectations based on past achievement of the control and experimental groups.

The Programs

A TEMAC program on positive and negative numbers of first year algebra, containing 500 frames, was used in a class of 34 Grade 10 boys, average age 15½, taking B.C. Mathematics 21. The other, a Grolier Society program on ratio, proportion, and variation, containing 360 frames, was used in a class of 29 Grade 11 boys, average age 16½, taking B.C. Mathematics 31. Both these classes are vocational, containing students of below-average academic ability.

The Operation

Both classes used the program in about the same manner. The purpose of the experiment, the theory underlying programming, and a little about their own particular program and its *modus operandi* were briefly explained to the class. The students worked individually on their own after that until they were finished. The teacher spent the period checking their work to ensure they were doing it properly and answering the individual questions. The unit on signed numbers took seven periods, the one on ratio and proportion, eight periods. The actual time spent working on the program was an average of about 30 minutes out of 44-minute periods. The remaining time was spent in class discussion of the program or in reviewing answers to short homework assignments in arithmetic fundamentals given most nights throughout this time.

After two days, it was obvious that their rates of progress through the booklets were greatly varied and I allowed the slower ones to take their programs home in order to keep the class together to a degree. Some students finished three days before the rest of the class even then. The program on ratio and proportion included a few test questions scattered throughout which the students brought to the teacher to mark. This had the advantage not only of testing the student with an unanswered question frequently, but also of keeping the teacher in touch with his progress. At the completion of the program each class was given a short examination two days later. In order better to test the effectiveness of this kind of instruction, no extra teaching was done before testing.

Results

The examinations given to these two classes were also given to other control vocational boys' classes of similar ability taught in the usual way by classroom teachers. To give a comparison of average abilities, the results of the Mathematics 21 and Mathematics 31 general Christmas examinations are included in the table below. The teachers of the *control* classes involved made no special effort to cover the material of the program but taught the subject as it naturally occurred in their syllabus. In general, the length of time spent on the topic by the control classes was slightly less than the time required by the class on programmed learning.

As can be seen from Table C-1, the Grade 10 programmed instruction group did not learn as much as the teacher-taught classes, even though they demonstrated a slightly superior ability on the Christmas examination of the full term's work. The Grade 11 class, however, did as well as the others, although their Christmas examination results indicated they should have done even better.

Because of the brevity of the experiment, the inexperience of class and teacher, the fact that the classes compared were not especially matched, and other reasons, these results are far from conclusive. The nature of the vocational boys classes should also be considered. University program students might have done much better.

Table C-1.—Comparison of Classes Taught by Teachers with Classes Taught by Programmed Learning

Class	Results of Examinations on Topic Taught by Programmed Learning to Show *Comparative Achievement of Classes*			Results of Christmas Examination on Full Term's Work to Show *Relative Abilities of Classes*		
	High	Average	Low	High	Average	Low
Grade 10:						
Positive and Negative Numbers						
Experimental Class	20	9.3	4	95	62.0	17
Control Class A	19	10.0	2	92	54.7	14
Control Class B	19	13.7	6	83	57.1	18
Control Class C	18	10.2	2	84	58.7	14
Grade 11:						
Ratio, Proportion and Variation						
Experimental Class	29	19.5	5	73	54.5	29
Control Class A	25	16.5	6	65	42.0	22
Control Class B	29	19.5	3	69	47.5	13

Opinions of Students and Teachers

Of course, opinions differ, but the following represents a cross-section of opinions held by the students and the teacher:

1. Programmed learning is a less interesting form of learning. Perhaps with more experience by both class and teacher this would be changed.
2. Knowledge gained from programmed learning has been thoroughly learned and may be retained for a longer time.
3. Poor students still have difficulty. One boy frankly admitted cheating on the program.
4. The program on ratio, proportion and variation seemed considerably better than the one on signed numbers. Students complained that early questions on the latter were too trivial and yet they had difficulty with the later ones. However, much of this might be attributed to their own attitude.
5. Programmed learning would be excellent for individual remedial work.
6. Programmed learning was not a teacher-saving device in this experiment. Individual assistance and general supervision took nearly all the teacher's time in the period.
7. The major advantage of programmed learning is that it enables a student to proceed at his own pace. However, in our present public educational system where students are examined periodically together, promoted at regular intervals, and generally kept in groups, programmed learning is less flexible than our present methods, and hence probably less useful. In a school where a student could proceed at his own pace permanently without being required to have covered the same body of information as other students have in the same time, programmed learning has real possibilities. This summarizes my present opinion; however, I must emphasize that our short use of it in two vocational boys' classes was far from conclusive. I still feel very open-minded towards the topic.

D. SIGNED NUMBERS: AN EXPERIMENT IN PROGRAMMED INSTRUCTION

by

DONALD M. RICHARDS
Rochester School
Rochester, Alta.

Fifty-three Grade 9 pupils from three schools used programmed instruction experimentally. An equal number from two schools receiving teacher instruction formed a control group. Four tests of intelligence, reading ability and arithmetic achievement were given to all pupils. It was concluded that the two forms of instruction were equally effective and that programmed instruction provided for individual differences in pupils better than teacher instruction.

The Programs

The program used in this project dealt with positive and negative numbers in algebra. Since this topic forms part of the Grade 9 mathematics curriculum in Alberta, it was possible to use the programs experimentally and to compare results using this method with results obtained by classes receiving teacher instruction as part of the regular mathematics course.

The Groups

Four schools offered to let their students take part in this project. In two of the schools, all Grade 9 students used programmed instruction. In another school, the Grade 9 students receiving teacher instruction as usual, were designated as control classes. In the fourth school, one class used the program while another did not. With this division, there were 53 students in the total experimental group and an equal number acting as controls.

The Project

Each of the groups was given four tests; a reading test (*The Gates Reading Survey*), an intelligence test (*Otis Gamma*), a test of arithmetic ability (teacher-made), and an achievement test on signed numbers (teacher-made). The intelligence test, reading test, and test of arithmetic ability were given before the start of the program. The test of achievement was given after the completion of the program. The tests of arithmetic ability and achievement in signed numbers will be referred to as the pre-test and post-test respectively.

The reliability of the standardized tests used in the statistical analysis was that reported by the publishers. The teacher-made pre — and post-tests had reliabilities of .75 and .61 respectively (Kuder-Richardson 21). These reliabilities tend to limit the conclusiveness of the results.

Hypotheses

Anumber of hypotheses were stated about the effectiveness of programmed instruction. It was hoped that this study would prove or disprove these hypotheses. Two hypotheses that were tested are as follows:

1. Short programs are as effective as teacher instruction.
2. Programs give more allowance for invidual differences than do teachers.

Observations

The standard deviations of the two groups are significantly different on the pre-test of arithmetic ability. Thus, the experimental group is more heterogeneous than the control group. However, after the experimental group had used the programs, the difference in heterogeneity between the two groups was no longer statistically significant. No significant difference in achievement was observed between the two groups on the post-test of arithmetic achievement.

The scores on the various tests are shown in Table D-1.

Table D-1. Means and Deviations of Two Groups Receiving Either Programmed or Teacher Instruction

Test		Control	Experimental
Intelligence	Mean	105.6	105.0
	Standard Deviation	11.6	10.7
Reading	Mean	69.2	66.9
	Standard Deviation	22.3	20.0
Pre-test	Mean	32.8	32.4
	Standard Deviation	7.2	8.5
Post-test	Mean	16.2	15.2
	Standard Deviation	4.5	4.3

Conclusions

The results of the experiment as shown in Table D-1 were interpreted as follows:

Hypothesis 1: Short programs are as effective as a teacher. This hypothesis was proven to be true. The means for the two groups on the post-test were not significantly different.

Hypothesis 2: Programs give more allowance for individual differences than do teachers. As shown by the standard deviations of the two groups on the arithmetic ability test, the experimental group is more heterogeneous than the control group. On the standard deviation of the post-test, there is no significant difference so the program has made them more homogeneous. In the process, it is likely that the better students same down slightly while the poorer students came up slightly.

E. THE EFFECTIVENESS OF PROGRAMMED LEARNING MATERIALS

by

E. PRICE and E. DE PAOLI
McKim Junior Secondary School
Kimberley, B.C.

Sixty-six paired students matched for intelligence and pre-test achievement formed two groups using as aids to instruction in mathematics either programmed materials or an authorized textbook. The work of both groups was also supplemented by teacher-presented materials. After 13 hours of instruction, no difference in achievement were observed between the two groups. Without the supplementary teacher instruction, the group using programs fell behind the teacher-instructed group. It was concluded that both methods of instruction were equally effective, although by implication programs were best suited to a role as teaching aids rather than independent teaching devices.

Object

To compare the effectiveness as teaching aids of the textbook prescribed for Grade 8 mathematics in British Columbia schools (Brumfiel, Eischolz, Shanke, *Introduction to Mathematics,* Addison-Wesley Publishing Co.) and of programmed learning materials (Temac Programmed Learning Materials, "Sets, Inequalities, and Functions", and "Positive and Negative Numbers", of *First Year Algebra,* Encyclopaedia Britannica Press Inc.)

Background

For the purpose of this experiment programmed learning materials were regarded as *teaching aids.* Set theory is taught in British Columbia schools as a means of developing concepts of algebra, and negative numbers are defined in terms of the basic properties of a number field. The Temac programs teach sets as a topic distinct from the algebra of real numbers and approach the study of negative numbers in a traditional manner.

On the other hand, much of the set notation used in the Temac program is normally introduced in Grade 9 mathematics in British Columbia schools. It was necessary, therefore, to supplement both types of materials in order to make a valid comparison.

General Experimental Procedure

Sixty-six Grade 8 pupils were divided into 33 matched pairs. The test group of 33 pupils (Group A) were then taught two units of work using the Temac programs supplemented by teacher-presented material, and the control group of 33 pupils (Group B) were taught the same two units of work using the prescribed textbook supplemented by teacher-presented material.

Pupil pairs were matched on the basis of five test scores. Three of these scores were obtained from a Pupil Aptitude-Achievement Report made to the school by the Division of Tests and Standards of the British Columbia Department of Education, after a provincial testing of Grade 7 pupils in March 1963. They consisted of an IQ score (Lorge-Thorndike — reduced to a ten-point scale), an arithmetic fundamentals score, and an arithmetic reasoning score (both from a *Metropolitan Achievement Test Battery* — both reduced to a ten-point scale). The other two scores were the results of a test of material presented to the pupils immediately prior to the experiment and of a pre-test of material to be taught during the experiment.

A comparative analysis of pupil scores is provided in Table E-1.

In addition to ability and achievement, sex and allocation of pupils to classrooms were considered in selecting pairs. Twenty-four pairs are matched according to sex and 24 pairs are matched according to division. Table E-2 gives sex-division comparisons of the experimental groups.

Table E-1. Median Scores of Two Groups of Thirty-Three Students on Five Tests

	Group A	Group B
IQ — Lorge-Thorndike	5	4
Arithmetic Fundamentals — Metropolitan	5	5
Arithmetic Reasoning — Metropolitan	5	4
Test of Previous Work	20	21
Pre-test of Experimental Work	7	5
Total	40	40

Table E-2. Other Considerations in Selecting Pairs

	No. in Group A	No. in Group B
Classroom 8	17	16
Classroom 9	16	17
Total	33	33
Boys	18	17
Girls	15	16
Total	33	33

Teaching and Testing Procedure

The experiment required 25 hours of class time (including teaching and testing).

During the first ten hours, Group A worked through the two programs while Group B worked through Chapter 16 (Sets and Variables) of the prescribed textbook. It was necessary to supplement the material of the textbook with more advanced notation. At the end of ten hours, the pupils of Group B wrote Test II (Solution Sets) and the groups changed teachers.

During the next ten hours, Group B received instruction in Negative Numbers (Chapter 18 of the prescribed textbook).

As the pupils of Group A completed the programs, they were asked to rewrite Test I (Knowledge of Experimental Work) and then given supplementary instruction in applying set theory to the algebra of real numbers. During the twentieth hour the pupils of Group A wrote Test II (Solution Sets) and the pupils of Group B wrote Test III (Negative Numbers). At this point the groups again changed teachers.

To complete the experiment, Group A received three hours instruction in applications of negative numbers and wrote Test III (Negative Numbers), while Group B reviewed previous work and received supplementary instruction in handling cartesian products. During the 25th hour Test I (Knowledge of Experimental Work) was administered to all pupils.

In order to eliminate, as much as possible, differences in instructional methods and in the marking of tests, all pupils received instruction in a given topic from the same teacher, and all tests which were used for any comparison were marked by the same teacher.

No reference has been made to "time spent on the program", although this point is stressed in most reports on programmed learning. In this experiment it was assumed that those who finished most quickly would spend more time on other work than those who finished more slowly, and therefore that the former would show their superiority, if any existed, in the results of their final examination.

Results and Comparison of Tests

The results of the tests, comparing the two methods of instruction, are given in Table E-3. For each test, median scores for the two groups are reported. Tests of significant differences between the groups were made using the deviations in scores of the paired students. The group using programmed instruction wrote Test I on three occasions — first as a pre-test, second after completion of the two programs and finally as a post-test following teacher instruction. The results of these tests are given in Table E-4.

Table E-3. Median Scores of Two Groups Receiving Programmed and Non-Programmed Instruction

	Test I Knowledge of Experimental Work			Test II	TestIII
	Pre-Test	Post-Test	Gain Pre- to Post-Test	Solution Sets	Negative Numbers
Group A (Programmed Instruction)	7	28	20	8	45
Group B (Non-Programmed Instruction)	5	28	21	7	48

Table E-4. Median Scores on Test I by the Group Using Programmed Instruction

	Test I	Test I	Test I
	Pre-Test	After Programs Completed	Programs + Teacher Instruction
Median	7	15	28
Gain		+ 8	+ 13

Observations

1. No statistically significant differences between the two groups were observed in their gain scores (Test I), in the test of Solutions to Sets (Test II), or in the test of Negative Numbers (Test III).
2. Group A pupils made greater improvement with teacher instruction than by working through the program alone.
3. Analysis of scores revealed no correlation between the difference in test scores for the two groups and sex, IQ, classroom, or pre-test results.

Objective Conclusions

Under the conditions of this experiment, little or no difference was observed between the effectiveness as teaching aids of Temac programmed learning materials and the prescribed textbook for Grade 8 mathematics in British Columbia schools.

Subjective Conclusions

1. The comparison of the scores of Group A on Test I at three times during the period of instruction suggests that the experimenters were correct in assuming that programmed learning materials should be regarded as teaching aids.
2. The subject matter and approach of the Temac programs are not consistent with those of Mathematics 8 in the program of studies for British Columbia schools.
3. The pupils of Group A become bored after working on the programs for a few days. Since the material tested by Test II was not covered by the programs, it is felt that the score in favour of Group A may be due to the fact that the pupils, having tired of the programs, were extremely attentive when they returned to normal teaching methods. It is further felt that the score in favour of programs indicates that the pupils of Group A were less attentive when material covered by the programs was extended.
4. Pupils who finished the programs most quickly scored no higher than their matched pairs in Group B, despite efforts by both teachers to take advantage of the time thus gained. This might indicate that pupils profit less than is commonly believed from being allowed to progress at their own rate.

Although the experimenters made every effort to remain as objective as possible, both felt that programmed learning *should* provide a very efficient method of allowing for differences in learning rates and every effort was made to ensure that the time gained by the faster students would be used to extend and complete their knowledge of the two topics. Although it is conceded that this experiment proves very little, it was nonetheless surprising to find no significant differences between the scores of the faster students of Group A and their counterparts of Group B.

Evaluation of the Experiment

Although an experiment such as this proves little about the effectiveness of programmed learning in general, it allowed the teachers concerned to examine one type of programmed learning materials and revealed certain weaknesses in that material.

As the results of the experiment were counter-balanced by previous successful experience with teacher-written programs, the overall experience has led to the belief that short, teacher-written programs will prove more effective than long, commercially-produced programs. It is certain however, that any program must be tested and judged on its own merits.

It is felt especially that the investigation of any program or teaching method is very worthwhile despite the results and even despite possible limitations in the validity of the results.

F. AN EVALUATION OF PROGRAMMED INSTRUCTION IN CHEMISTRY USING MATCHED PAIRS OF STUDENTS IN GRADE 10 AND GRADE 11

by

DOUGLAS K. JARDINE
St. Patrick High School
Arvida, Quebec

Seventeen matched pairs in Grade 10 and fourteen in Grade 11 were subjects in the experiment. Two groups in each grade received either programmed or teacher instruction in chemistry. Pre- and post-tests were given to the Grade 11 students but only post-tests to those in Grade 10. No significant differences were observed between pre-test scores of the experimental and control groups in Grade 11 nor between their post-test scores, although both groups improved significantly during two and one-half weeks of instruction. Similarly, few differences arose between post-test scores of the two groups in Grade 10. It was concluded that pupils learned equally well under both kinds of instruction.

General Information

This experiment in the use of programmed teaching was carried out in St. Patrick High School, Arvida, Quebec, during the month of May, just before the final examinations for the 1963-64 school year. The high school has an approximate enrolment of 180: 81 are in the senior high school consisting of Grades 10 and 11. Although the school functions as an English-speaking one, the French language is more truly the vernacular of the student. The school offers only one set of courses; those leading to university entrance. The students' backgrounds are varied, but many have parents who are professionals in the employ of The Aluminum Company of Canada, or Price Brothers. About 10 per cent of the pupils enrolled have marked difficulty with the academic courses offered in the school. For example, there are 17- and 18-year olds enrolled in Grade 9 and 20-year olds in Grade 11.

Programmed Material Used

The programmed materials used are listed in Table F-1.

Table F-1. Programs Used in Test

Name	Used In Grades
1. Chemical Symbols	10 and 11
2. Valence	10 and 11
3. Formulae and Nomenclature of Binary Compounds	11
4. Formulae and Nomenclature of Acids and Salts	11

Since machines were not available, the programs were used with paper to "hide the answers". Students were asked not to peek, and by and large they refrained from so doing. They were also asked to record their answers on paper and these were later scanned to help gauge the application of the student.

Group Employed in Tryout

The experiment was carried out in the subject of chemistry, and in particular the areas listed in Table F-1. There were essentially two experimental groups, Grade 10 and Grade 11. The average age of the Grade 10 group was 16 years and eight months, with an average IQ of 103. The average age of the Grade 11 group was 17 years and six months, with an average IQ of 107. These data are reported in more detail in Tables F-2 and F3. All students were part of the experiment, but only data from matched pairs were used in the statistical analysis. In Grade 10, 17 matched pairs were formed, and 14 in Grade 11.

Procedure Used in Tryout

In Grade 11, the programmed material was used as review in these topics for the provincial examinations. In the case of Grade 10 this was not the first introduction to the topic, but it was the first time that it was done in such detail.

The programs were used during the regular class time. Those using the programmed material were together in one room, but separate from those who were reviewing the same material with the teacher. A few students asked that they be allowed to take the program home, and they were allowed to do so. These students were in the main good but slow workers.

The students using the programs were allowed to ask the teacher questions on the material, but in actual fact very few questions were asked. This was true of both grades and was probably due in part to the subject matter being not entirely new. The students reported that they found the programs very straightforward.

No changes in this procedure were made during the course of the tryout, which spanned a period of two and one-half weeks — about ten teaching periods.

Learning Comparisons Made

Matched Pairs

There are two Grade 10's and two Grade 11's, the division being made more or less on proven ability. If a student had less than an overall 60 per cent average on quarterly examinations, he was placed in the "poorer" class. This division was not permanent but was reviewed after each set of quarterly examinations. There was no difference in the material taught or in the examinations written because of this division.

For the purpose of this tryout, the two classes of Grade 11 were considered as one, and matched pairs based on sex, age and IQ (but not on proven ability) were formed. This meant that pairs were not necessarily within one class and as a result members of such a pair would not work on the same material concurrently. Those who were not part of a pair were distributed between the test and control groups. The students were not told who was paired with whom or scores made on the tests, but were given all other details.

One half of these pairs were assigned to the control group and the others to the test group. The control group followed instruction with the teacher based on the material covered in the program. The time spent on each program was about a period and one-half, that is to say, 70 minutes. The test group set the pace and the control group followed. When a program was completed, a test on that material was given before starting on the next program.

Exactly the same thing was done with Grade 10.

Tests Used

Four tests — one for each program used — were prepared by the writer after he had examined the programs. Without forewarning, these tests were given to the Grade 11 students and these results used to establish their level of knowledge. No explanations or answers to the test questions were given at this time. These same tests were given again, one by one as each program was completed. These results, compared with the previous ones, gave the increase in knowledge.

In the case of Grade 10, no pre-program test was given. Only two tests were administered, although most of the test group covered all four of the programs. This was due to unforseen complications which arose at the time.

Results

A comparison of the results appears in Tables F-2 and F-3. The results for Grade 10 show a difference in favour of the experimental group; however, statistical analysis shows the difference to be not significant (Table F-2).

For Grade 11, the pre-program test averages show differences in favour of the control group, but are generally close to one another. This is not totally unexpected since the control group had more "better" students than the test group. The post-program test scores indicate about a 20-point increase as well as the same differences and closeness of averages. No test of significance was applied to either pre-program test and control comparisons or to post-program test and control comparisons, the reason being that the writer felt there was no significant difference — a feeling base on the Grade 10 results.

However, a test of significance was applied to results obtained before and after each program was used in both test and control groups. The results of these comparisons showed that the post-program results for both the test and control groups were significantly higher than the pre-program scores.

The only clear conclusion that can be made then is that learning was achieved in both groups, the method of instruction being immaterial in this tryout. Further, this effect was true for both grades.

Student Reaction

Student reaction was one of curiosity and increased effort irrespective of their group or grade. This decreased with time, fastest in the students of proven poorer ability. This was most true for the Grade 11 test group, less true for the Grade 10 test group and least true for the control groups.

Problems or Difficulties in Tryout

There were no real difficulties encountered in this tryout. In ordinary use, rate of progress would not matter — the faster students getting ahead and speeds. In the tryout it was necessary to brake the fast ones a little, and prod the slowest ones.

Some of the poorer students tended to approach the programs with less application than the others. In the final analysis, this is why they are classed as "poorer" students; not because of less ability. For these, the program was no more stimulating than the traditional method.

Comments

The author must admit to being pleasantly surprised with the results — results that don't really show up in numbers and which require a personal knowledge of the student to be really appreciated. Were I to be teaching next year, I would re-do this work, and definitely use programmed material for certain pupils at certain times.

When considering the rightful place of programmed teaching the writer feels you touch upon a huge philosophical problem. In brief, it should not be used as the main tool of education. Its efficacy will depend on the judgment of the teacher: a bright student can profit, a poor student can profit, but there are bright and poor students who cannot — who might instead profit from skillful manipulation by the teacher, if they profit at all.

Thought on this topic cannot assume that any one program will ever be the overall ultimate. In any one area, valence let us say, the writer feels that the choice of program will have to be based on the teacher's knowledge of the student: use of programmed material has always to be on an individual basis.

While programmed instruction can deliver ideas and concepts, the writer has a sneaking suspicion that it cannot imbue a "feel for the subject" into the student. This latter is more easily transmitted (not in fact an easy thing at all) with the flexible presentation feasible with a teacher.

Table F-2. A Comparison of Measures — Grade 10

	Program		No Program	
Average age — boys	17 yr. 1 month		17 yr. 1 month	
Average age — girls	16 yr. 0 months		16 yr. 5 months	
Average age — group	16 yr. 7 months		16 yr. 9 months	
Average IQ — boys	101		101	
Average IQ — girls	106		106	
Average IQ — group	103		104	
Total Scores — Average				
	Chem. Sym.	Valence	Chem. Sym.	Valence
Boys	70.6	71.1	67.0	66.3
Girls	88.0	72.1	80.4	69.0
Group	78.8	71.5	73.3	67.7

Table F-3. A Comparison of Measures — Grade 11

	Program	No Program
Average age — boys	17 yr. 11 months	17 yr. 7 months
Average age — girls	17 yr. 4 months	17 yr. 4 months
Average age — group	17 yr. 6 months	17 yr. 5 months
Average IQ — boys	106	103
Average IQ — girls	110	108
Average IQ — group	108	106

Test Scores — Average

	Chemical Symbols		Valencies		Binary Compounds		Tern. Acids		Chemical Symbols		Valencies		Binary Compounds		Tern. Acids	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Boys	58.2	76.5	34.8	42.4	50.6	59.4	39.6	60.4	62.9	79.3	74.0	76.6	67.0	74.2	53.7	62.9
Girls	64.3	84.7	52.3	66.3	54.8	76.4	37.9	71.8	66.5	86.7	49.6	61.6	60.4	77.9	43.3	69.5
Group	61.9	82.2	45.6	48.9	53.2	65.3	38.6	68.3	65.3	84.1	57.8	73.4	62.6	76.6	46.7	67.1

Significance: Total Group Before and After

0.01 level	0.05 level	0.01 level	0.01 level	0.01 level	0.01 level	0.01 level	0.01 level

G. PROGRAMMED INSTRUCTION AND TEACHER INSTRUCTION: A COMPARISON BETWEEN STUDENTS IN GRADE SEVEN AND GRADE EIGHT

by
MARSHALL P. BYE
Standard School
Standard, Alberta

Twenty pupils in Grade 7 and sixteen in Grade 8 were matched on the basis of past achievement in mathematics. Two groups in each grade received either programmed or teacher instruction on the topic of ratio, proportion, and variation. Comparisons of achievement between the groups were made after ten periods of instruction. Conflicting and inconclusive results were obtained from these comparisons although programmed instruction was never superior to teacher instruction and was sometimes inferior in effectiveness. The study also examined the effectiveness of the programs with pupils of higher and lower intelligence and of those with satisfactory or unsatisfactory work habits.

Problems for Investigation

1. Do students do as well on programmed instruction as they do under teacher instruction?
2. Do certain students do better on programmed instruction than they do under teacher instruction, and vice versa?

Design: Hypothesis

Students receiving instruction from a teacher do better on tests than students using programmed instruction.

Pupils in Grades 7 and 8 were paired according to past performance in mathematics. One member of each group worked with programmed material (Group A) while the other member received regular classroom instruction (Group B). Achievements of these two groups were compared in total and for each grade. The effectiveness of the program was also studied with pupils of high and low intelligence and for those with good and bad work habits.

The program used was "Ratio, Proportion, and Variation" from TMI - Grolier.

One teacher, following the program content very closely, taught Group B with normal classroom procedures, but without using textbooks for instruction. Duplicated pages of assignments and notes were given from time to time. During this instruction, the teacher assigned exercises to be completed in class time, made use of blackboard drill and gave individual assistance when required. No spot tests were administered.

A second teacher was in charge of programmed instruction for Group A. Close supervision was maintained to discourage cheating. Occasionally, this teacher supplemented the programmed instruction by teaching a regular lesson when required by the group. Other activities included individual assistance, blackboard drill, and spot tests. She put students back in the course to repeat when results indicated this was necessary.

Both groups spent ten class periods of a half hour each on the material. No time outside of class was spent on the work. When both groups had finished the material a test was administered and marked. This test was made up by both teachers working in close cooperation.

Comparisons of Achievement by the Groups

Statistical tests were applied to the results to make comparisons between the effectiveness of the two kinds of instruction with the various groups. The tests used were the sign test and the "t" test. Using the sign test, the following results were obtained:

1. No group achieved better results from programmed instruction than under teacher instruction. ("Better" in this sense means that the results on the final tests under one method were significantly higher than under the other method.)
2. For the two classes combined, teacher instruction was better than programmed instruction.
3. For the Grade 7 class alone, the same results were achieved as for the total group. The group receiving teacher instruction was superior to the group using programmed instruction.
4. For the Grade 8 class alone, students did as well on the programmed material as under the teacher instruction.

Using the "t" test for difference between means, the following results were obtained:

1. No group did better on programmed instruction than under teacher instruction.
2. The combined classes did as well on programmed instruction as under teacher instruction.
3. The Grade 7 class learned better under teacher instruction than on programmed instruction.
4. The Grade 8 class did as well on programmed instruction as under teacher instruction.

The achievement of the pupils was also compared on the basis of intelligence and work habits. Two groups were identified according to intelligence with those of IQ greater than 110 in one group and those under that point in the other. The choice of 110 as a cutting point was arbitrary. Results obtained were as follows:

1. In Grade 7, there were ten pairs of students

matched by IQ. Of these, six had IQ of 110 or over. Of this more intelligent group, the average score for students receiving teacher instruction was 22.1 per cent higher than the score for those using programmed instruction. Four pairs of students had IQ's under 110. In these cases, the average score for students using programmed instruction was 3.3 per cent higher than the score for those receiving teacher instruction.

2. In Grade 8 the students, when matched according to IQ, had almost the same scores on the test.

To compare the effect of the two kinds of instruction on pupils of different work habits, the following comparisons were made. The students were grouped, on the basis of teacher opinion, into those having good work habits, and those having poor habits.

1. Of ten pairs in Grade 7, all five good students did better on teacher instruction, while four of the five poor students did better on teacher instruction than under programmed instruction.
2. Of eight pairs in Grade 8, all good students (four) did better on teacher instruction than under programmed instruction. Of the poor group, three out of the four did better on the programmed instruction than on the teacher instruction. One did better on the teacher instruction.

Conclusions

1. In Grade 7, teacher instruction produced better results than did programmed instruction.
2. In Grade 8, there was no significant difference in the results produced under the two methods of instruction.

Comments

1. We failed to give a pre-test, and therefore were unable to measure the growth or amount of learning. Grade 8 students had been exposed to some of the work earlier, and were therefore not particularly attentive nor interested during teacher instruction. Under programmed instruction, the novelty of the presentation maintained interest in the topic.
2. On the use of programmed instruction the instructions to the students were:
   - (i) On first question wrong, go back five questions and repeat these.
   - (ii) On second question wrong within five questions of the first question wrong, go back ten and repeat.
   - (iii) If student still has difficulty, ask assistance from the teacher.

Observations

1. The teacher had to constantly be putting students back. Many students hurried through, many cheated, many asked no questions even when there was difficulty, many wouldn't go back when they should (on their own), many gave little attention to the routine practice and explanations.
2. Even with a small class of 16 pupils, the teacher did not have enough time to give the assistance that was required. It was frustrating work for the teacher in that often she had to teach some things 16 times.
3. Some parts of the programmed course were not well developed and clearly needed teacher introduction. This supplementary instruction was given in this experiment and hence the results of the groups using programmed materials may have been higher than would be obtained without teacher instruction.
4. Some students finished early. Some students never did finish. The first three "finished" had the lowest scores on the test.
5. The difference between the time required for good students to finish and that required by poor students seemed to be increased by the use of the programmed materials. The teacher felt that slow students would be slower still and good students would progress more rapidly than under traditional teaching.
6. We restricted the class period to one-half hour. Our normal mathematics period is one hour. From earlier experience with programmed instruction, we had found that in periods of much more than one-half hour the students tended to drop in efficiency and in interest. With this in mind, we restricted the length of each period.
7. Plans are being formulated to carry out the same project on a larger scale next school term by enlisting the cooperation of neighbouring schools. Plans also are being made to determine the effect of the length of a class period on programmed instruction results.

H. AN EXPERIMENT IN THE USE OF PROGRAMMED LEARNING

by
D. L. FEIR
Grand Forks Secondary School
Grand Forks, B.C.

Two heterogeneous classes of Grade 9 pupils received either programmed or regular instruction. Working without any direct teaching, the programmed instruction group completed their work more quickly than the control group but obtained lower results on a final test.

The experimental work for this report was carried out in the Grand Forks Secondary School in February 1964. Testing was completed February 18.

The purpose of this experiment was to compare the learning accomplished through a program with that learned from a teacher using orthodox classroom methods.

A program consisting of 360 frames on "Ratio, Proportion, and Variation", published by the Grolier Society, was supplied by the Canadian Teachers' Federation. This program is a unit from *Algebra II* prepared by this company.

Grade 9 pupils were selected because they had had no direct contact with this subject previously and because the work seemed useful to our classes at this age level. Sixty children were divided into two heterogeneous classes so that they would be as nearly equal as possible. Division was made on three bases:

1. sex
2. scholastic ability
3. achievement on a mathematics examination given in December 1963.

The classes were assigned to a teacher by tossing a coin.

Class A was given the program with enough instruction to teach the pupils how to use it. They received no assistance with the subject material. The average time required to complete the work was four hours. It varied from 3 1/2 to 4 1/2 hours.

Class B received instruction from a teacher. The lessons were prepared from the material in the program so that identical information was presented to both groups. This required seven hours of teaching.

At the end of the teaching session both classes were tested at the same time. The scores, converted to percentages, are given in Table H-1.

Table H-1. Comparison of Results

Class	Median	Time Used
A Programmed Instruction	46.5	4 hours
B Teacher Instruction	57.5	7 hours

The conclusion suggested by these results is that a class may be taught in less time by a program but will not learn as thoroughly. However, this is not likely a sound conclusion since the teacher might well be able to cover the subject material of the program in an equivalent time. His natural tendency is to drill or re-teach when difficulties appear.

Pupil reaction to the program varied. Some were bored; some liked to work alone; others thought that a combination of programmed and teacher instruction would be excellent.

I. PROGRAMMED INSTRUCTION AS REVIEW FOR PUPILS IN HIGH SCHOOL CHEMISTRY

by
G. J. DICKSON
Charles Bloom Secondary School
Lumby, B.C.

Teacher instruction in chemistry was given to 24 pupils in Grades 11 and 12. Following this instruction, pupils used either programmed materials or class notes as review. It was concluded that the programmed review of instruction in chemistry was more effective than teacher review or review by pupils using their own notes.

This is a small secondary school enrolling some 180 pupils in Grades 8 to 12. It is located in a small logging-farming community some 16 miles from the nearest population centre.

The class using the programs consisted of 24 pupils — thirteen in Grade 11 and eleven in Grade 12.[1] Of these, in Grade 11, four were boys, nine were girls; in Grade 12 four were boys, seven were girls. Data relating to the achievement of these pupils are given in Table I-1.

Table I-1. Average Intelligence and Academic Ability of Pupils

	Intelligence	Achievement: First Quarter	Achievement: Second Quarter
Total Group	115.3	68.0	75.6
Group A	114.5	69.7	75.1
Group B	116.0	64.1	76.2

Students in both grades take the same science course with chemistry and physics being given in alternate years.

IQ's were determined by means of the *Otis SA Tests of Mental Ability High Form C.* Course achievement represents the two report card grades given in November and February. Each of these is arrived at by averaging a number of hourly tests given throughout the quarter with a two-hour comprehensive examination given at the end of the quarter. These are raw scores, in this instance, and were not scaled.

For the first part of the experiment the class was divided in two by random selection. Group A was given the *Valence* and Group B the *Symbols* program. Both of these topics were familiar to the pupils from instruction in Grade 10 and from application in their current courses. Pupils completed one or other of the two programs according to their assignment to a group. In this way, pupils completed programmed review on either valence or symbols but not on both topics. No review was given by the teacher.

Each pupil completed the program on his own and recorded the time spent. When the programs were completed, the class wrote tests on valence and symbols.

The second part of the experiment consisted of teaching *Binary Compounds* to the whole class, then giving the programs to Group B as review. Group A used, as review, notes taken during the lessons. After the programs had been completed, a test on the material was administered.

Part three consisted of teaching *Acids and Salts* to the whole class, giving the program to Group A and allowing Group B to use notes only for purposes of review. After the programs had been completed, the teacher did a complete review with the whole class and then administered a pertinent test.

Some of the control effect may have been removed during this third part of the series by switching programs between the two halves of the class. The switch was made for two reasons: (1) Both programs were complementary in the nature of the material covered and therefore, if there were benefit to be gained, the teacher felt it to be unjust to deprive half the class of it; (2) If the superiority of either method of review were to be established, it would be better shown by comparison with a greater number of results.

A small modification in the material covered was made in this section. The program contained no reference to di- and tri- basic acids. The teacher included these as a more logical approach to the naming of acid salts than the method used. A suggestion to this effect was forwarded to the CTF Research Division for inclusion in revision of the program.

The results obtained by using the two methods of review are presented in Table I-2.

Table I-2. Average Achievement Scores for Pupils Using Two Methods of Review in Chemistry

Test	Method of Review	Score
Test 1 Symbols	Programmed Review (Group B)	80.9
	No Review (Group B)	25.1
Test 1 Valence	Programme Review (Group A)	80.0
	No Review (Group A)	75.0
Test 2 Binary Compounds	Programmed Review (Group B)	95.2
	Note Review (Group A)	75.3
Test 3 Acids and Salts	Programmed and Teacher Review (Group A)	93.4
	Note and Teacher Review (Group B)	78.6

From the results of the series the following conclusions appear to be obvious:

1. Some review is better than none at all.
2. Programmed review is very superior to note review.
3. Programmed review is slightly superior to teacher review.

Other conclusions not so obvious, perhaps, may also be drawn:

1. IQ does not appear to play a significant role. Those students whose overall performance was high on the term work scored well regardless of their preparation for the test.
2. Generally, scores for the second quarter, during which the experiment took place, were higher than those of the first. Part of this may be ascribed to more familiarity with the course.
3. The relatively minor differences between the scores on the *Symbols* and on the *Valence* programs may be ascribed to the continuous use to which this knowledge, gained in Grade 10 and used throughout the course, is put. The other parts of the experiment involved new cr almost new concepts.

J. PROGRAMMED LEARNING AS AN INSTRUCTION AID IN ALGEBRA

by
J. F. CLARK
Langley High School
Langley, B.C.

This report makes a critical examination of programmed instruction in Grade 9 algebra. The program used was the Temac series entitled *Algebra I* which closely parallels the Mathematics 10 course of the British Columbia Curriculum. The programs were used by students throughout most of the algebra sections of the course during several months of the school year. Precisely matched or equivalent groups were not available for experimental comparison with the group using programs. However, several other teacher-instructed classes were tested and compared to the class using the programs. The study revealed many problems associated with using programmed materials over a lengthy period for instruction. It suggests too, some issues which must be resolved if the full potential of programmed instruction is to be developed.

Objectives of the Evaluation

The advantages claimed for programmed learning include the following expectations:

(a) The provision for individual differences through varying rates of progress.

(b) Superior motivation because of constant knowledge of achievement.

(c) Opportunity for enrichment because the teacher is freed from the usual preparation of lessons and lecturing.

(d) These advantages result in a better achievement score for the majority of pupils.

Since arrangements to use the program were not made until the last week of August, no attempt could be made to assign pupils to classes so as to form matched or equivalent groups for experimental purposes. As a result, one of several existing classes was selected to use the program simply because it was a large class (39 pupils) and because the least number of pupils in it were enrolled in other courses taught by the teacher handling the programs. A suggestion was made and considered which would have meant that the pupils would receive all instruction in algebra solely from the text. This proposal was rejected since the programmed material was considered to be an aid to teaching rather than an independent and sufficient method of instruction. The study was therefore primarily concerned to discover what problems the method would either solve or create in a normal school setting. In short, the conditions under which the field test was carried out were not planned for rigorous experimental research purposes. No hypotheses were tested and the only concession to formal research procedures involved a quickly organized standardized testing program to supplement the usual teacher-made tests. In effect, we were interested in trying the programmed materials to determine their effectiveness and to determine the reaction of pupils and teachers to this new form of learning.

Grouping Procedure

After seven weeks of school during which all six classes taking the Mathematics 10 course worked on the units in construction geometry, we were ready to begin the Temac program. The last week of October and early November were spent in administering three sets of standardized exams to each of the six classes — about 180 pupils. The results are summarized in Table J-1.

Table J-1. Pre-test Results of Six Grade 9 Mathematics Classes

Class	N	Otis Beta IQ Standard Deviation	Otis Beta IQ Median	Otis Beta IQ Mean	Iowa Aptitude Percentile Standard Deviation	Iowa Aptitude Percentile Median	Iowa Aptitude Percentile Mean	Cooperative Math Standard Deviation	Cooperative Math Median	Cooperative Math Mean
Experimental	39	7.3	111	111	15.2	71	71	20.2	70	69
C	39	8.4	113	111	15.9	78	74	16.1	82	77
D	27	8.7	107	108	15.2	72	70	19.0	78	73
E	23	6.9	111	110	16.2	66	63	18.3	68	63
A	31	7.3	104	104	18.6	69	63	21.9	63	61
G	29	9.6	107	107	17.8	77	72	24.2	82	72
All Classes	186	8.6	109	108	17.1	72	70	20.8	74	69

It was hoped that this assortment of marks could be compared with similar measures at the end of the experiment. Such factors as the resignation and reassignment of a teacher resulted in classes A and G being exposed to three different teachers during the year. Two of these were in their first year of teaching the Mathematics 10 course. Coordination of work and the labour involved in handling nearly 500 tests such as these on an extra-curricular and do-it yourself basis resulted in some loss of enthusiasm on the part of the author who, in addition to his other teaching duties, was director of the project.

Some Observations

The most important observation that can be made is not peculiar to programmed learning. It is that any radical change in teaching procedures must involve a great deal of teacher time. In addition to the work involved in attempting a comprehensive testing program, such unusual items as a paper shortage, a stencil shortage, a shortage of qualified typists, not to mention the breakdown of a privately owned typewriter with an engineering keyboard, all contributed rather more than their share of headaches. As a result, great difficulty was experienced in producing sufficient unit tests to enable one to keep track of individual progress. This testing aspect of programmed learning is extremely important. The small-step feature of the Temac programs demands that frequent test situations be made available in order to reassure the pupil that he is making progress. The tests also provide a very rapid assessment of the areas of difficulty and provide the teacher with the knowledge needed to help each pupil as necessary. If students have difficulties at some point in the program, it may be necessary to supplement the program with teacher instruction. This need becomes apparent when pupils suddenly find their test results plummeting.

Many of the problems of evaluating pupil progress would not be present after the initial year's work. Because of them, most class time was spent working with individual pupils. No real attempt to provide the hoped-for enrichment was undertaken. The classroom was packed with pupils. There was no space to have small groups of pupils working on any of the enrichment materials available. A certain amount of interference was the inevitable result of explaining and discussing areas of difficulty with the pupils on an individual basis.

The problem faced by the teacher in answering pupil questions is severely emphasized with programmed materials. Some pupils appear to ask questions whenever they want a break — others who should, never do. A rather high percentage of questions are simply time-killing devices used by the pupil who is too lazy to read intelligently — at other times, real problems exist. The teacher should know his class.

Pupil reaction is a function of several variables. It is essential to realize that not every able pupil wants to learn. The built-in motivating ideas in programmed learning will not work for every pupil. One of the definite drawbacks to the Temac courses used is their utter sameness from frame 1 to frame 8,312. Nowhere is there a chapter heading, a paragraph of summary material, an introduction to new units. This even flow of print takes no cognizance of the fact that learning proceeds by leaps and bounds — and has its levelling-off periods, too. Despite the rather negative views presented, it was clearly evident that an unusual majority of the pupils were willing to learn algebra from the text. Attitude tests were given at intervals. The final test of attitude was administered at the end of March and showed that only one of the 37 pupils was deeply dissatisfied with programmed instruction. A comparison with the attitude of adults to the same program was possible since a class of adults used the same program as part of a night school course in algebra. Their reaction was highly favourable and in many cases widly enthusiastic. They had a far greater appreciation of the advantages of being able to make progress on their own and to my surprise felt very definitely that they understood far more about mathematics than their previous school experience had indicated was possible. It is only fair to point out that no attitude test was given following the Easter exams. The results of this might have reduced the favourable pupil reaction considerably.

The problem faced in attempting to create a valid test which would be fair to pupils using either the regular text or the Temac text was further compounded by the tremendous variation in rates of progress by the pupils in the experimental class. In order to comply with the customary administrative arrangements in the schools, a common test was produced and used. The results were inconclusive on several counts. Certain areas in the Temac are not covered in depth — the loss here is made up in breadth of topics. Reconciliation of these differences and an attempt to be just in terms of amount covered by the conscientious members of the experimental class resulted in a difficult problem in evaluation to which there was no satisfactory single solution. Since all pupils were measured against a common standard, some injustice was bound to occur. It would seem to be essential to realize that if the content of a text is appropriate to the aims of a course, then the content of that text should be used as a basis for evaluation. The impossibility of being fair to two dissimilar texts (in a single trial at least) should have been obvious.

Final External Test

The final test selected to evaluate the programmed instruction was the Cooperative Algebra Test (Form Y). It was selected before the instruction began because it appeared to provide sufficient coverage for pupils who might finish the Temac program. (The last 2,500 of the total of 8,312 frames were at a Grade 10 mathematics level.) Unfortunately, none of the pupils finished the Temac course. As a result, the standard deviations for each class were almost as great as the median scores of the classes. (See table J-2). That is, the marks were spread in a highly asymmetrical fashion about the central scores with low scores predominating despite a number of very high marks. We could state that the Temac group achieved a median score about two points above the average of all other classes (which simply reflects the results of similar initial standard tests), but in view of the overall failure of the test to give reasonable scores, no conclusions can be drawn from the quantitative evaluation of the program's effectiveness.

Table J-2. Final Test Results of Six Grade 9 Mathematics Classes

	Percentile Scores			
Class	Standard Deviation	Median	Mean	N
Experimental	24.4	29.8	33.8	36
C	22.5	36.5	38.2	37
D	22.4	24.7	25.9	24
E	15.2	14.9	20.1	18
A	20.8	16.5	25.5	26
G	24.1	38.5	42.5	28
Total (Non-Experimental)	23.3	27.4	32.0	133

Attention should be drawn to the fact that class G (which has much the highest median scores) was the one previously referred to as having three teachers. Is this the answer to our problem?

Recommendations

1. Teachers intending to use programmed materials should be familiar with the background and theoretical aspects of the development.
2. Initial use of programs (particularly if extensive standardized testing is planned) should be accompanied by some extra help in handling the additional work required.
3. Since the use of programmed texts is likely to be rather unusual for some time, provision should be made for pupils to elect out of the class. The unwilling pupil is probably somewhat safer in a programmed learning class than under more normal situations.
4. An attempt to cover a limited amount of work, say one or two units of a course, is likely to provide a more reasonable basis for evaluation than a long-term effort such as was carried out in this study.
5. The implied benefits of providing for individual differences by widely varying rates of coverage might, in the initial trials, be sacrificed by definite and generally achievable assignments being made. Rapid workers could be given enrichment activities or alternate texts. The teacher would not lose control of the teaching situation. Introductory material, review discussion and common tests could all be part of the process. These features are not possible to the same extent if every pupil is on his own for any length of time.
6. Teachers using programmed learning materials must be familiar with their subject matter. A class of 40 pupils can ask a surprising number of varied and unrelated questions in an hour period. One doesn't have time to do basic subject matter research in the classroom.
7. The reaction of pupils will be determined to a great extent by their knowledge of progress — this has implications for *any* teaching situation — therefore great care must be exercised in test construction and grading.
8. Pupils (as well as teachers) must know what difficulties to expect when using programmed learning texts for the first time. Programs definitely demand an unaccustomed maturity and sense of responsibility on the part of the pupil. He must realize that the ultimate pay-off is going to be determined by the conscious desire he has to use the material properly.
9. Certain problems in this evaluation may be traced to the fact that only one teacher in the school was primarily involved with the use of this material. However, a great deal of interest in the project did develop in the local area and throughout the province, after the project was under way. However, it could have been a great advantage to be able to test the materials thoroughly in the school and to "shake down" ideas and problems on the job before exposing them to public examination and criticism.
10. As a result of many discussions with teachers throughout the province, it is felt that realization of the full potential of programmed instruction will never be achieved in the existing educational framework. Programmed instruction would appear to fit in rather well with the concept of team teaching in which time for independent study is specifically allowed. In our system, its chief use would seem to lie in the enrichment area or in enabling small schools to provide a broader offering (if suitable programs exist). It has definite possibilities for private study by adults and in correspondence courses, even in its present state of development.
11. A final important implication derived from the use of programmed instruction should be mentioned. The characteristic of this instructional medium — especially its impersonal aspects — may be expected to have some effect upon the quality of thinking which it induces in the learners who use it. Our ideas about evaluation, enrichment, acceleration and methods of teaching and about learning theory and the ultimate objectives of education will need to be re-examined and brought into sharper focus, if such means of instruction are to be developed further in our schools. If enough people can be critically involved with such new instructional devices as programs and educational television, they will have served their purpose regardless of the ultimate success of the "gadgets" themselves.

K. AN EXPERIMENT IN THE USE OF PROGRAMMED MATERIALS

by

L. W. ROBINSON
Geraldton Composite High School
Geraldton, Ontario

Nineteen students in a control group were matched on the bases of age, sex, and intelligence with 19 students in an experimental group. Tests of algebra achievement, problem solving, critical thinking, and study habits were administered as pre-tests and their alternate forms as post-tests. Two forms of a test of attitude to mathematics were also given to the groups. The experimental group used a programmed text in Algebra from October to May while the control group received regular classroom instruction. A teacher-made test of achievement in algebra showed no significant difference between the groups after instruction by the two methods. However, comparison of gains by the groups on a standardized test of algebra achievement showed a significantly greater gain for the experimental group.

Location of the Investigation

In September, 1962, the Geraldton Board of Education, Geraldton, Ontario, purchased thirty sets of Temac *First Year Algebra* for experimental use in a Grade 9 class of the Arts and Science Division. The investigation subsequently conducted at the Geraldton Composite High School is the subject of the following report.

Reason for the Investigation

The general purpose of the study was to observe the actual use of programmed material in a classroom situation, and from such observation to gain some insight into the proper utilization of this new technique. In theory, as well as in specific applied instances, programmed material appeared to have shown itself to be a successful instructional method. The author wished to discover whether a comparable level of success could be attained in a school area where students were drawn from an essentially working class background. To facilitate the study the author enrolled in a Programmed Learning Seminar conducted at Guelph Agricultural College in the summer of 1962.

Design of the Study

The Population and Sample

The Geraldton Composite High School is situated in northern Ontario, approximately 170 miles northeast of Port Arthur. It has a population of 450 students, half of whom come from the neighbouring communities of Longlac, Beardmore, Jellicoe, Nakina, Caramat and other outlying districts. There are five classes at the Grade 9 level.

The students registering in the five-year Arts and Science Division were paired as closely as possible on the basis of sex, chronological age, and I.Q. These students were then randomly assigned to either the control or experimental class. The pairing had to be done on the basis of information available from the various elementary schools and from option sheets completed by the Grade 8 students. Because the I.Q. scores reported were from several different tests, the Dominion Intermediate was administered in late October, 1962, in order to verify the scores used for matching. Although 24 students were originally placed in each group, some difficulties were encountered in September when a few students chose to transfer, or dropped out. In effect, the matching provided for two types of comparison:

(a) Employing 19 pairs of students matched closely on intelligence, age,and sex.

(b) Employing two groups of 23 and 24 students respectively, who formed the units of instruction for which the average values on the matching variables were closely equated, and for which the effects of non-matched variables should have been more or less equalized by randomization.

There was, however, a general feeling among the staff that the experimental group was a stronger class intellectually than the control group. While it cannot be described statistically, this strength and enthusiasm did, nevertheless, appear to exist.

The control group met from 9.00 to 9.35 a.m. and the experimental class from 10.10 to 10.45 a.m., five days a week. Both classes met in the same room, selected the same obligatory and optional subjects, and had the same teacher for each individual subject. The author was in charge of both mathematics classes.

Tests Selected for Criteria and Control Measures

The following tests, because of their availability, low cost, and ease of administration and scoring, were selected for criteria and control measures. The first five tests shown in Figure K-1 have parallel forms. The pre-experiment and post-experiment scores are therefore available. The tests were administered in mid-October, before the work with programmed material began.

1. Algebra Fundamentals (Australian Council for Educational Research)
2. Problem Solving Tests (Arithmetical and Algebraic)
3. Critical Thinking (Watson-Glaser)
4. Inventory of Study Habits (Holzman-Brown)
5. Attitude Toward Mathematics (teacher-made)
6. Teacher examination in Algebra
7. I.Q. Test (Dominion Intermediate)

Figure K-1. Criteria and Control Tests Employed.

Conduct of the Experiment

The experiment with the programmed material began on October 23, 1962, at which time the work covering experimental geometry had been completed in both classes. This period allowed sufficient time for the students' initial adjustment to high school and to establish rapport between the students and instructor.

In the experimental group the programmed material was introduced as simply another method of learning mathematics. The fundamental rules for using the program were explained and the first few questions were completed with the students working as a group. Tests were then prepared by the instructor so that the student would write a one-period test approximately every 500 frames. In the class following the test, the student received a remedial program based on the results of his test. If the student completed all questions correctly he was allowed to proceed with his program with no remedial work. For all other students, depending on the nature and number of their errors, the remedial program consisted of a verbal explanation, re-doing appropriate frames and completing questions from their standard textbook. It was carefully explained to the students that these tests were not used for promotion purpose but rather as a technique to evaluate where they needed additional practice. Since all students wrote the same test, but at varying times, it would have been a relatively simple procedure for the slower students to determine the content of the test. This, however, did not appear to become a problem.

From this testing it became quite apparent that some difficulties were not being uncovered in sufficient time to make remedial work feasible. Moreover, the "form" in which answers were required was causing considerable difficulty. It was, therefore, found necessary to introduce more written work and at frame 2099, and at numerous subsequent frames thereafter, the students were required to complete selected questions from their standard text at th black-boards. This procedure also allowed time for intelligent discussion with the students when their black-board work was corrected.

Many students expressed some difficulty at working on their programmes for 35 minutes at one sitting. It was therefore considered sensible to break up some of these periods with meaningful work. The students' routine in class then began to include working from their programs, related board work, testing, remedial work and discussion.

Students were permitted to take their programs home but no compulsion was placed upon them to do so. At various stages, however, students were made aware of approximately how many frames should have already been completed if they wished to finish their program in the school year, and some of the slower students were occasionally encouraged to show more industry. In the control group, homework initially was not assigned (to control the time variable), but questions were suggested for review and additional practice. As this procedure appeared to have a detrimental effect on their attitude, homework was re-instituted after approximately 3 weeks.

All students finished the program by May 25, 1963; the quickest moving student was finished on May 13, the same time the control group completed its prescribed course.

Each student was required to keep a list of the actual frames completed each day, and whether these frames were completed at home or in the class period.

In examining the Temac program, it was discovered that it contained frames that were not applicable to the Ontario Grade 9 course of studies and these (Table K-1) were omitted entirely.

Table K-1. Frames Omitted

FRAMES
1514 - 1970
3655 - 4318
5287 - 5369
5421 - 5461
6542 - 7551
7971 - 8015
8129 to end

Analysis of Data

Method of Analysis

Initially, averages were computed for all control and criterion measures for the two entire groups. Tests of statistical significance were applied to the 19 matched pairs only, because the statistical status of the complete groups was somewhat cloudy. In any case, our computations indicated that the results obtained with the matched pairs would be nearly identical with those obtained had the entire groups been used.

The t-test was applied to:

(a) all differences between means of the experimental and control groups on all criterion and control variables

(b) differences in the gains on the criterion variables between the experimental and control groups.

Finally, a graphical analysis of the inter-relationships of several variables was undertaken to shed further light on the relationships involved.

Additional Analyses

In addition, extensive analyses were undertaken of the number of frames completed each day and evening by each student in the experimental group. The results for the three repeating students in each class were examined in detail. Moreover, a chi-squared analysis of changes in student attitudes towards programmed instruction between December and June was completed.

Results of Analysis

Background Data

Table K-2 reveals that there were no significant differences between the means of the experimental and control groups in the control variables, with the exception of reasoning. The difference in reasoning was significant at the .01 level in October but was not significant at the .05 level in June. Moreover, the two groups were closely equated in another standardized and highly regarded test of reasoning ability (Watson-Glaser). Consequently, one may be justified in regarding the difference as either a statistical long shot (which could be attributed to testing variation) or as the effect of some unknown extraneous factor which was operating at the time the test was administered.

Table K-2. **Background Data (Averages) on 19 Matched Pairs**

Variable	Experimental Group	Control Group	t-Test
Age (Sept. '62)	14.64	14.61	0.15
I.Q. (Oct. '62)	108.7	108.5	0.15
Sex - Males	9	9	—
- Females	10	10	—
Reasoning	11.6	9.2	3.39
Holzman-Brown	38.8	41.6	.94
Australian	17.4	19.8	1.03
Watson-Glaser	68.1	69.7	.57

Table K-3. Criterion Data (Averages) for 19 Matched Pairs

Variable	Experimental Group	Control Group	t-Test
Christmas Algebra	16.5	18.0	0.68
Christmas Geometry	40.8	37.3	0.92
Australian Algebra (June)	116.1	103.8	1.71
Final Algebra	47.1	43.2	0.68
Gain in Australian Algebra	98.7	84.0	2.06

Criterion Scores

In the teacher-made final algebra examination the observed differences in means were not significant at the .10 level but were close to significance (Table K-3). This particular examination (of one hour's duration) was prepared with a great deal of care in order that it would reflect the terminal behavior prescribed by the Ontario course of studies. The examination paralleled those of previous years so that its bias, if any, would favour the control group.

Again, on the June form of the Australian test, the t-test favoured the experimental group, but the observed differences in means were not significant at the .10 level. The *gain* in performance on the same test, however, was significant at the .05 level. As this test measures the capability of students to deal with the basic operations in algebra, the results on this particular test are extremely important.

Table K-3 shows that at Christmas the control group performed slightly better than the experimental students in algebra and slightly worse in geometry. This fact seems compatible with the gains in performance on the Australian algebra test wherein the experimental group gained a greater mastery of fundamental operations, and outperformed the control group on all criterion tests.

Figure K-2, exhibiting the relationship between IQ and gain in Australian algebra, shows a higher trend line in favour of the experimental group. Figure K3, showing the relationship between IQ and achievement in algebra, illustrates a greater correlation between programmed instruction and IQ than between traditional instruction and IQ. In interpreting this graph, however, several pertinent facts should be kept in mind. In the control group, those with an IQ below 90 were repeaters and hence they tended to score slightly higher than those with an average IQ of 100. For the programmed group, the repeating students fell into the group with an average IQ of 100 and hence moved the curve upward in this area. The shape of the curve in this area, then, is due to the small sample and the difference in IQ's of the repeating students rather than some innate characteristic of the program. In the upper areas of the curve, the experimental students clearly outperformed the control group. This is probably due to "teaching to the middle of the class" and, to some extent, ignoring both the gifted and the extremely slow.

There was no appreciable difference between the programmed and control groups in English and geography, but there were significant differences in favour of the experimental group in French composition, history, music and typing. The Holtzman-Brown *Survey of Study Habits* showed no appreciable difference in *change in* study habits, although both groups exhibited a significant decline in study habits.

It would be extremely dificult to attribute this decline to any specific factor, but most certainly it did not seem to be influenced by one form of instruction as opposed to another.

Attitude Change Analysis

The attitude survey taken in January and again in June showed an overall *change* in favour of traditional instruction that was "significant" in every sense of that word.

Ten of the 24 students agreed, however, that they would like to learn other subjects by programmed instruction. Fourteen students felt that programming was a boring way to learn, but as one student noted on his questionnaire, "some teachers can be boring too". It seems noteworthy that in many cases it was the better student who rejected programming. There appear to be two reasons for this rejection. In the first place, the gifted student was compelled to complete more frames than he felt were necessary to learn a particular point. A second reason, which was implied but not stated, was that students were forced to be mentally active for an entire period whereas in an ordinary classroom situation much of their time could be spent on their private thoughts while basic work was being drummed into the slower students. While this observation does not reflect favourably on the industry of some students it appears, nevertheless, to be a legitimate one.

The students did not sign their names to the questionnaire and hence felt more free to record their genuine feelings than they might otherwise have been.

Interpretation and Conclusions

The Christmas geometry work, which favoured (slightly) the experimental group, might have indicated that the experimental group would also outperform the control group in algebra, had they continued to be taught by their instructor. This, coupled with the fact that there were significant differences (or differences approaching significance) in favour of the experimental group in French composition, history, music and typing (subjects which were not programmed), together with the general staff feeling that the experimental group was a slightly stronger class, contributes to a rather confused picture. While it is true that the experimental group outperformed the control group, one is left to ponder what might have occurred had both groups beer taught by traditional methods even though, statistically, there were no significant differences in control variables. Nevertheless, it is the opinion of the author that if the times had been equated and if both groups had approached the learning task with equal zeal, the experimental group would still have outperformed the control class. This opinion is based in part upon subjective evaluations and in part upon the realization that more time was consumed by testing and related board work (which for some students was not always necessary or beneficial), in the experimental class than in the control group.

The overall trend toward a dislike of programmed instruction would certainly have been more pronounced had the experimenter not included board work and discussion. Many students simply found the program boring. This was especially true for the gifted students who were forced to wade through frames they found relatively simple. This difficulty, however, could be overcome through the use of a branching rather than a linear program, or at least a program directed at a specific intellectual level. Perhaps, instead of *a* program, there should be three or four programs, each aimed at differing abilities. It is no more reasonable to expect one program to be suited to all intellectual levels than it is to expect one teaching method to be equally suited to the gifted and the retarded.

Programs of a unit design would be more economical and flexible than the present Temac format. If smaller units, perhaps two weeks' work, were available, an instructor could adapt the unit either to introduction, presentation or review purposes, whichever best suited his needs.

The unit design would also overcome the tendency for students to look at the answers before they have attempted the question. This problem does not manifest itself until the students are well advanced with the program. An individual who finds himself considerably behind his classmates (either from lack of industry, or genuine difficulty), and who is embarrassed by the fact, increases his speed by looking at the answer, *then* writing it down.

The program did not result in any time-saving for the students in this experiment. Unfortunately, the students, without exception, did not view the program as an opportunity to progress more quickly and proceed to the work of the next grade. As previously mentioned, the students seem conditioned by our educational system, wherein one does sufficient work to stay out of trouble and obtain "good" marks but does not work hard enough to make himself conspicuous by being too far ahead. Had the programmed instruction experiment been carried out in Grade 13, where there exists the pressure of an external examination, I believe one would have noticed a considerable decrease in the time taken. Moreover, the older the student, the further removed he is from the idea that the teacher is solely responsible for his progress and that it is the teacher's task to ensure that he passes. Programming demands maturity on the part of the student; the maturity to work on his own initiative, the maturity not to look at answers, and the maturity to recognize that he is responsible for his own progress. Perhaps a typical Grade 9 student is *not* sufficiently mature to benefit from sustained work on a program, but needs the enthusiasm and encouragement which is, hopefully, derived from a human teacher.

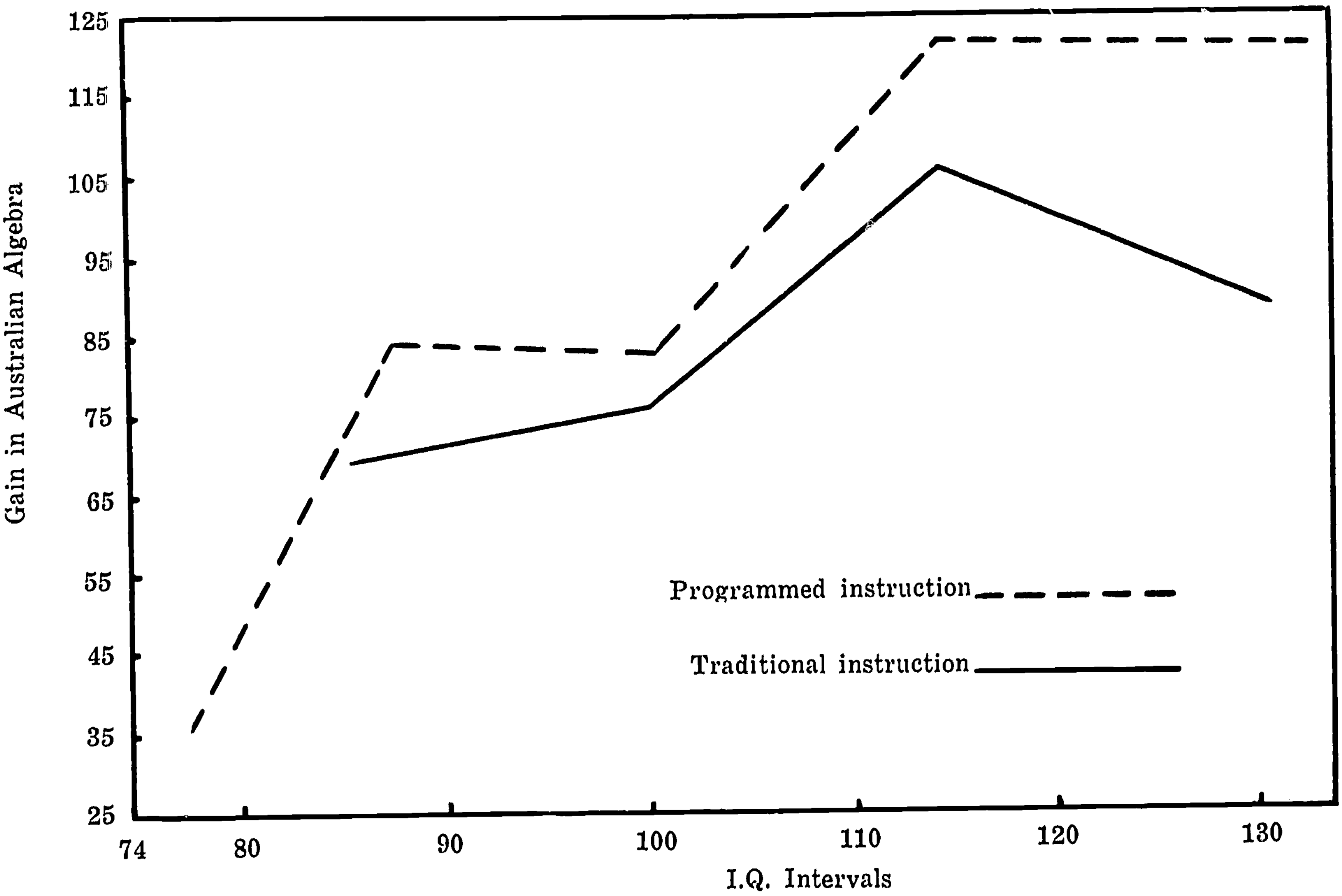

Figure K-2. Relationship of I.Q. and Achievement Gain in Australian Algebra for Programmed Instruction and Traditional Instruction.

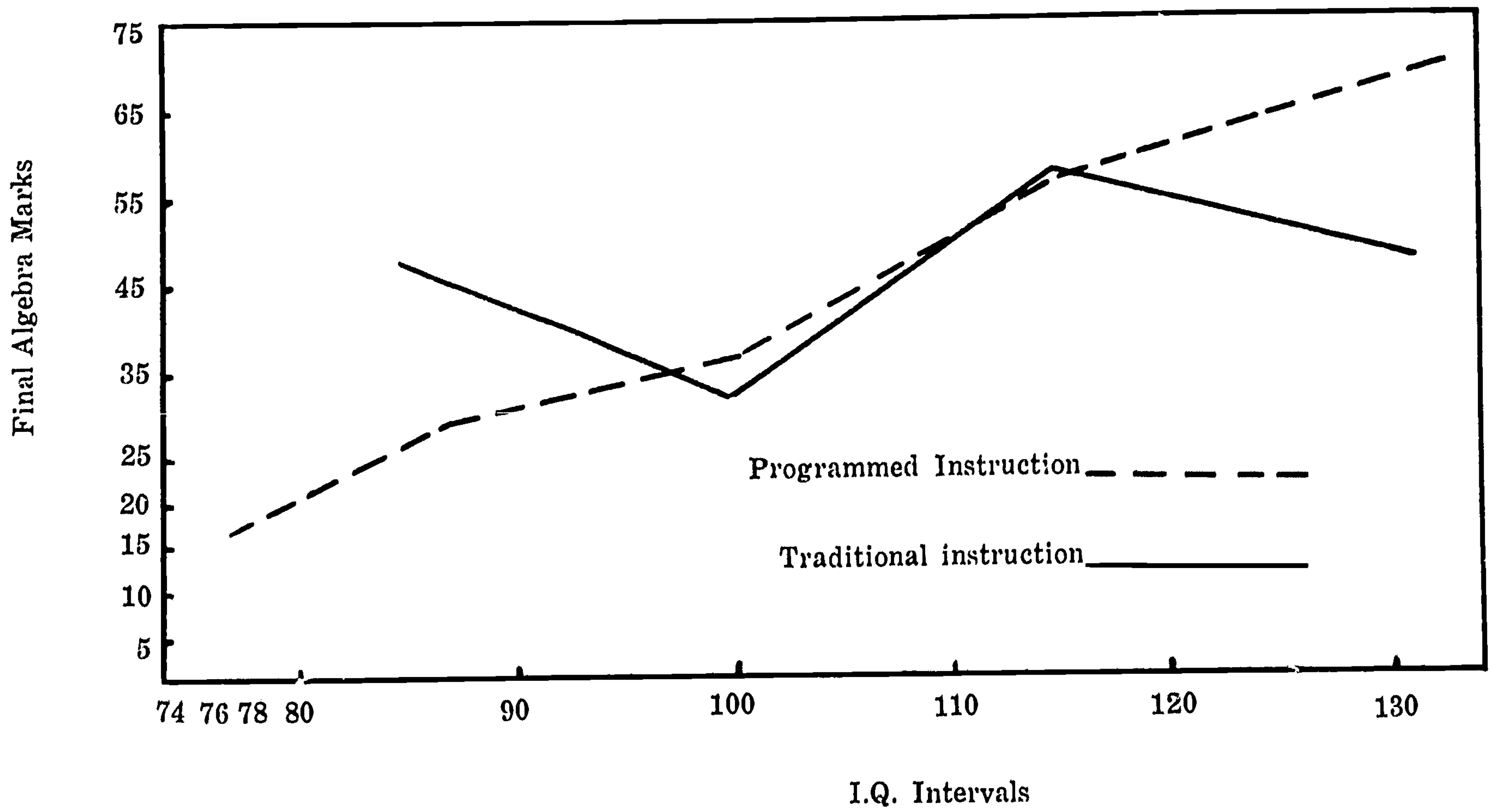

Figure K-3. Relationship of I.Q. and Achievement in Algebra for Programmed Instruction and Traditional Instruction.

CHAPTER THREE

ORGANIZATION

OF

TEACHER RESEARCH

CHAPTER THREE

ORGANIZATION OF TEACHER RESEARCH

For teacher research to be purposeful and meaningful, a certain climate conducive to scientific investigation must prevail in schools. Moreover, the teacher researcher must possess a number of skills and have access to the resources which make controlled observation possible. What specifically are these conditions that are necessary or desirable for the production of worthwhile teacher research? Several considerations should be mentioned.

Conditions for Quality in Teacher Research

The first requirement for classroom research is that the teacher understand the logic of controlled experimentation. Clearly the experimental method is not equivalent to research itself and it is therefore desirable that the teacher be familiar with general research issues and specifically with other research methods. However, as in other areas of scientific enquiry, experimental procedures are of overriding importance in classroom research. Although experimental design can reach levels of formidable complexity that are beyond feasibility in the classroom, the basic requirements of comparison between matched groups are simple and readily adapted to the conditions which prevail in classrooms. By employing the logic of this method of enquiry, the teacher wields a powerful tool in research.

The question of what research designs are suitable for classroom use will be considered in a subsequent section of this report. The designs of the studies reported here will also be compared to these desirable features of classroom experimentation.

A second necessary condition for adequate classroom research into the problems of instruction requires that the teacher possess the techniques and understanding involved in applying tests of statistical significance to experimental data. Both statisticians and their textbooks sometimes manage to obscure or bury deep in too much explanation the profound but essentially simple ideas that can be of greatest use to the researcher. Teachers, school administrators and parents are used to dealing with measures of pupil achievement and ability as precise and reliable data for such purposes as the ranking of pupils, asignment to class or course, promotion, awarding scholarships and for making many other decisions of this kind. These decisions are so common and necessary as to be the categorical imperatives of education systems. Therefore, the statistical notions of "sample," and "significant difference" have to be introduced as a part of teacher research which bears little relation to the usual process of decision-making carried on with the normal data which schools collect about pupil achievement and ability. Fortunately, the idea of chance differences arising between the achievement of two comparable classes is not a concept which runs counter to common sense and the essentials of statistical decision-making should therefore be easily acquired by the teacher researcher.

Familiarity with certain psychometric ideas concerning the meaning of a test score is a desirable asset for the teacher researcher. Concepts in the reliability and validity of tests would assist the teacher to seek, and apply, tests which yield the greatest information and provide adequate bases for drawing conclusions and implications from observed results. Similarly, such knowledge and skills might reduce the urge to test and measure beyond the teacher's capability or resources for analysis. Teachers, like some other researchers, may yield to the temptation to measure and correlate everything in sight and still leave the essential problem of their study unanswered. Thus, it is to be hoped that familiarity with basic statistical and psychometric concepts can strengthen and give precision to teacher research in the classroom.

A third prerequisite for useful teacher research is a working knowledge of other research dealing with the problem which is to be investigated. Teachers are essentially subject-matter specialists and their training and daily experience is therefore centered around this specialty. When teachers do research, the topic or problem posed does not usually concern their specialty as such. Rather, the topics of teacher research often deal with novel methods or applications associated with the teacher's subject specialty. For example, most of the studies reported in this publication deal with the teaching of mathematics. But the research deals not with mathematics itself but with the effectiveness of programmed instruction in selected mathematical topics. Clearly then, a teacher's mathematical specialty does not necessarily prepare him to deal with programmed instruction. Therefore, before effective classroom research can be devised, the teacher must become familiar through study with the thought and research behind this new technique. Such study provides familiarity with the essential features of a procedural innovation or instructional device and, hopefully, suggests problems for research that are appropriate to the unique features of classrooms. In this way, teachers can avoid the unnecessary duplication of research and can develop research which leads towards the most promising areas for investigation.

A fourth requirement for quality research concerns the flexibility of the school organization within which the study is conducted. Some research in the classroom can be carried on without disturbing the usual routine, organization, scheduling, or course of study prevailing in the school. Experimental groups may be constituted from among existing classes or parts of them; the new procedure under evaluation may not require a change of course requirements; a teacher may wish only to alter experimentally his methods of procedure without altering the externally fixed aims of instruction. In many other instances, however, the research or evaluation requires modifications in the organization or administration of the school which are beyond the scope and authority of any teacher individually. For example, a decision to use televised instruction or team teaching, to evaluate a new curriculum, or to assess certain materials of instruction not normally supplied to teachers would require resources and administrative powers beyond those that are usually at a teacher's command. Moreover, the experimental trial of certain new curricula or courses of study may imply not only new methods to achieve old aims; it may imply as well the adoption of totally new aims and purposes. Thus, flexibility of school organization and program is frequently essential to implement teacher research projects having larger scope or purpose.

A fifth desirable condition in establishing teacher research is a liberal supply of and access to the resources which are necessary for the efficient pursuit of a research study once its purposes and design have been established. Among the desirable resources are time for analysis of data or clerical assistance in performing such analysis. Access to calculating machines or automatic data processing can speed much of the drudgery of analysis. Research does not relieve a teacher from the other duties and responsibilities of his position. His research must usually be built upon a schedule which is already crowded and onerous. Thus, resources of time or clerical assistance can often make possible studies which would otherwise be beyond feasibility or would place unreasonable demands on the time and energy of the teacher.

Research in education is much dependent, too, upon measurement by standardized tests. Such tests are available in varieties which can be suited to a multitude of purposes. Though they are expensive in relation to teacher-made tests they have obvious advantages. Provision should thus be made for supplying special tests for use in classroom research.

A final requirement for producing worthwhile teacher research is related to some sociological considerations of group values, norms, and sanctions. At present, the teacher researcher is often unique among his peers, at least in the setting where he works, teaches and has contact with other teachers, pupils, and parents. However numerous researchers may be at special conferences or workshops, in their schools or district they may represent only one of many kinds of professional interest and activity in education. Moreover, teachers are trained to engage in didactic pursuits which are ordinarily not associated with research. The introduction of a Doubting Thomas in the form of scientific enquiry into an activity which otherwise requires certainty, continuity and stability is a move which may meet or create resistance, uncertainty or anxiety on the part of other teachers, pupils, or parents. Thus, research in the classroom, if it is to be successful, must elicit the understanding and support of other teachers and school authorities. Projects which run counter to the powerful expectations of a teacher's colleagues and peers can cause stress in staff relations and jeopardize the value of the project. As stated in the opening section of this publication, change is in many ways the basic reality of our world. Research implies and facilitates change. The teacher interested in classroom research often perceives the need for change that is not apparent to others around him. It is essential, therefore, that persons conducting research in the classroom be aware of the implications which such study has for the values and behaviours of others. Successful classroom research requires, too, that care be taken to explain the purposes of the research and to communicate to others its findings and conclusions. Similarly, school administrators and teacher organizations must facilitate this process of communication about research and strive to establish those conditions most favourable to its growth. In this way, teacher research can make its most useful contribution to problems in the improvement of instruction.

Designs of Experimental Classroom Research

Experimental research is by no means the only kind of research apropriate to the classroom, as was argued in an earlier section of this report. However, experimental designs are perhaps most common among the varieties of research conducted by teachers in their classrooms. One of the reasons for this situation may be that the logic of controlled comparison has an esential simplicity which is easily adapted to the classroom. All the studies reported here are experimental in the sense that they incorporate some form of a design that compares two groups, one of which is subjected to all but one of the significant conditions which apply to the other. Thus, in these studies, there are typically two classes or groups, one of which receives usual teacher instruction while the other receives the experimental treatment. Such procedures by no means represent a solution to all problems in design of experiments but they represent the essential feature of the method — controlled comparison.

Commonly, however, school evaluations of new procedures, materials, or curricula are called "experimental" when in fact they lack any provisions for controlled observations. "Experimental" in this sense means embarking on a course which had not been followed previously in this particular school or classroom. The most severe limitation of this procedure as research is that it provides no means of discovering whether other classroom conditions would have produced the same results as the one actually employed. The experiment therefore lacks adequate control. In diagramatic form, the design of such experimentation is as Figure 1.

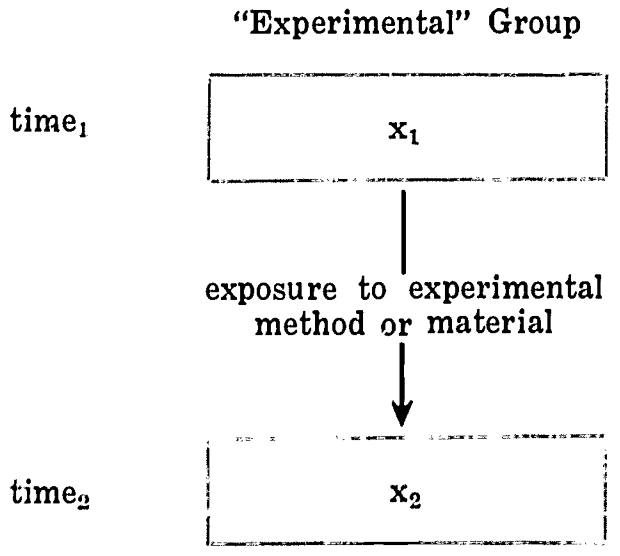

Figure 1. Experimentation Without Control.

For example, a teacher may wish to decide whether a procedure for teaching spelling in several short periods is superior to teaching the same material in a massed period of time. At time$_1$, therefore, the teacher gives his class a test in the material to be taught and proceeds to teach the asigned words during many small periods of instruction equal in total duration to the time he would spend under the previous method. At time$_2$, he again tests the pupils using the same test given at time$_1$. If x_1 and x_2 are the class averages at the two times, it would be safe to predict that the average mark received by pupils at time$_2$ would be greater than that at time$_1$. But is this result different from what might have been achieved under other teaching methods? The teacher has no way of answering such a question objectively although he will undoubtedly have formed some opinions about this matter.

Many research studies in school essentially follow this model of "experimentation" given in Figure 1. Unfortunately, it is never possible in social research to return to time$_1$ and repeat the experiment using another method of instruction. If such reversion were possible, as it usually is in the physical sciences, many of the difficulties of social research would be removed.

To overcome the problem outlined above, the method of controlled comparison is employed. A model for this design is given in Figure 2. As an example of

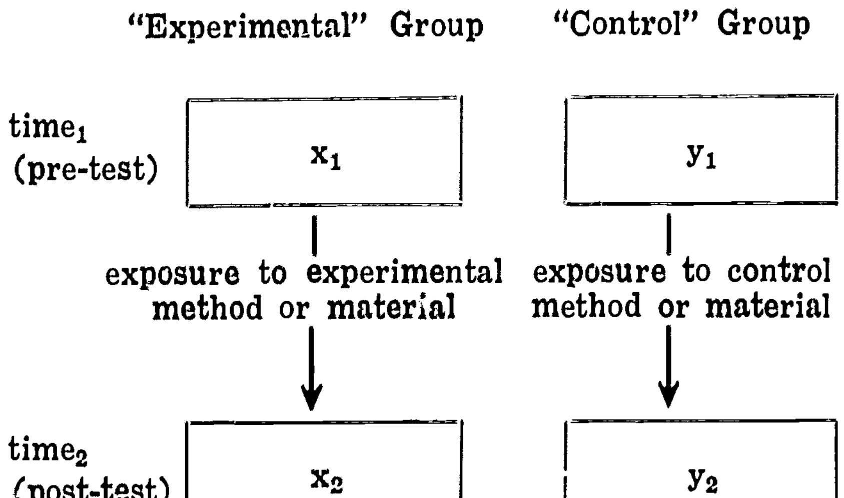

Figure 2. A Model for Controlled Experimentation.

the use of this model, a teacher might employ a new method of spelling instruction with the experimental group and his previous method with the control group. Both groups would be given identical spelling tests at time$_1$. Such tests serve as pre-tests of spelling ability and can be used to determine whether the groups were in fact equivalent before the experimental method is applied to one of them. Then x_1 is the average mark in the experimental group and y_1 the average mark in the control group at the beginning of the experiment. To make a comparison at a later time, these two groups should be equivalent in all significant respects at time$_1$ — i.e., the value of x_1 should be equal to y_1. In other words, each group should be able to spell equally well or badly at the beginning of the experiment. After the groups have received instruction by the two methods, the same test should be given as a post-test to the groups. The average scores are now called x_2 and y_2 and any significant differences between them could be attributed to the difference in the effectiveness of the methods used, if we can assume that the groups were in fact equivalent initially in all respects but the manner of instruction.

In algebraic terms, the effectiveness of the experimental method of teaching spelling could be expressed as follows:

$$\text{Effectiveness of the experimental procedure} = (x_2 - x_1) - (y_2 - y_1)$$

This equation implies that the effectiveness is the difference observed between the average marks of the two groups at time$_1$ and time$_2$. But if the groups were equivalent in spelling ability at time$_1$, then $x_1 = y_1$ and the equation is simplified as follows:

$$\text{Effectiveness of the experimental procedure} = x_2 - y_2$$

Thus, if two groups are equivalent at a given time, *any* test given at time$_2$ can be used to determine the effectiveness of instruction received between those two times. Therefore, a teacher might use one test to establish equivalent groups for experimental purposes and another to determine whether the experimental treatment produces significantly different results from the control method.

Establishing Equivalent Groups

An essential though often difficult task in completing an experimental design arises in establishing equivalent groups for comparison. Schools frequently assign pupils to classes on any of a number of administrative criteria. For example, they may be grouped according to achievement, by patterns of courses elected, or merely by alphabetical order. None of these methods can be expected to yield equivalent groups for experimental purposes, for each method of selection builds in a biased selection. For larger groups, a random selection is the most satisfactory basis for constituting the groups. That is to say, chance alone must determine whether an individual is placed in the experimental or control group. This kind of choice implies more than selecting a book "at random" from a shelf or a pupil name from a register since some purposive action is inevitably involved in such selection. Drawing names from a hat approaches a random procedure, though use of a table of random members is preferable.

Classes in school are usually too small to rely on randomization alone to produce equivalent classes. Therefore, some means of matching pupils is usually required. In matching, it is necessary first to decide what qualities the groups are to be matched on. For valid comparison, the groups should be matched on a characteristic which is relevant to the comparison which will be made after the experimental treatment. In the spelling research given as an example above, the pupils should be matched initially for spelling ability or achievement. Both groups should be able to spell equally well in the beginning since it is their spelling performance after the experimental procedure which is used to judge the effectiveness of the methods. The pupils could also be matched initially for age or intelligence. However, these characteristics of pupils are not necessarily closely related to spelling ability. Therefore, to match on either of these bases alone does not necessarily created equivalent groups.

If a group of pupils is to be matched for past achievement, the following procedures might be followed. Arrange the names of the boys in descending order according to their achievement on a standardized test of achievement in spelling. Select the first two names and toss a coin to assign one boy to the experimental group and the other to the control group. Proceed in this way with the remaining names on the list. Follow a similar procedure with a list containing the girls' names. The resulting groups will thus be matched on two criteria — sex and spelling achievement. Matching on more than two variables becomes progressively more difficult as more variables are added. Moreover, matching on further variables after the first few adds little more towards establishing the equivalence of the groups than matching on the initial variables. That is, having matched on such significant variables as, for example, intelligence, achievement, and sex, it is generally found that the groups are comparable also in other characteristics such as reading ability, mental age, chronological age and so forth.

If individual matching of pupils in the groups is not possible, then group matching may be used as an alternative method of establishing equivalent groups. Instead of pupil for pupil comparison, group measures such as the mean and standard deviation are considered. The aim is to create groups whose characteristics closely approximate each other. In this way, we may consider that the groups themselves are paired rather than particular individuals within the groups. If random assignment of pupils to groups does not produce equivalence, then individuals are dropped from each group until significant characteristics of the group closely approximate each other.

After randomization and matching of any kind, various measures should be used to check the equivalence of the groups. The hypothetical data given in Table 3 represent two groups of similar age, ability and achievement. If such groups had in fact been ectablished by random or matching procedures, a teacher could be assured of having achieved groups closely matched in these respects at least.

Table 3. An Example of Groups Matched for Experimental Purposes

	Group A		Group B	
	Mean	Variability	Mean	Variability
Intelligence test "X"	113.7	9.6	112.9	9.8
Age (in months)	188.3	5.1	185.8	5.2
Achievement test "A"	8.9	1.1	8.7	1.2
Spelling pre-test	9.5	1.1	9.6	1.0

These matched groups should then be assigned randomly (by flipping a coin, perhaps) to either the control or experimental procedures. Such groups would still not be perfectly matched, of course. Other unmatched variables including attitude to school, socio-economic background and group qualities such as sociability, and cooperative behavior could bias the outcome of the experiment. The assignment of teachers to the groups is another important variable which is usually non-random in educational research. With all these limitations, however, any random or matching procedures which establish equivalence between groups in some degree make important contributions towards controlled comparison and therefore towards the conditions necessary for experimental research.

Control and experimental groups need not be of identical sizes. When only limited numbers can be placed in the experimental group, it may be desirable to match two control pupils for each in the experimental group. In all cases, the essential criterion for the establishment of equivalent groups is randomness in the assignment of pupils to control and experimental groups. To accept pre-constituted groups for experimental groups is to run some risk of building in a bias which in the final comparison will favour one group or another. For example, two classes may be closely matched for intelligence, age and sex but one class contains pupils enrolled in academic courses and the other pupils in a vocational and technical course. If these two classes are then compared as to attitude to school, reading habits, or the effectiveness of different methods of instruction, a biased result will likely have been built in by the initial assignment of pupils to the two groups.

Some Further Design Problems

The experimental designs considered to this point have permitted only the very simplest kinds of comparisons, usually between two groups. In this system problems are stated to apply to the group as a whole rather than to parts of it. For example, we might ask which of two methods of instruction is more effective with groups of 20 to 30 pupils. Other kinds of problems might be posed as well. Do pupils of low and high intelligence learn more effectively with one method than the other? Is a particular method more effective with boys than girls? These questions cannot be answered reliably unless they are posed *before* the research is begun and unless a complex design is established to test for differences between groups of students. Suppose a question is posed as to whether programmed instruction is more effective with boys than girls. Instead of two comparison groups, four are now required, since the design incorporates the two dimensions of method of instruction and sex of pupils. The design for these groups is given in Figure 3, assuming that there are 40 boys and 40 girls in the total group.

Sex	Method of Instruction	
	Programmed	Teacher
Boys	20	20
Girls	20	20

Figure 3. A Two-Dimensional Design: Method of Instruction by Sex of Pupils.

Ideally, there would be four matched groups of 20 pupils each. For each type of instruction, there are two groups established according to sex of the pupil. Thus, boys using programmed instruction are compared to boys receiving teacher instruction. A similar comparison is possible for girls.

If another dimension such as intelligence is to be considered, the number of groups required is again doubled. The model for a three-dimensional design is given in Figure 4, assuming again that there are 80 pupils divided equally according to sex and intelligence groups.

Sex	Intelligence	Method of Instruction	
		Programmed	Teacher
Boys	High	10	10
	Low	10	10
Girls	High	10	10
	Low	10	10

Figure 4. A Three-Dimensional Design: Method of Instruction by Level of Intelligence and Sex of Pupils.

The problems of matching and randomizing are, of course, increased as the number of groups is increased and the complexity of the analysis grows correspondingly. The number of pupils in any one group in such designs is usually small since the number of pupils available is often restricted. Teachers, then, are usually compelled to limit their designs to simple comparisons involving usually no more than two groups. It is, of course, possible to perform a series of comparisons between groups taken two at a time, although the groups used in each experiment should be composed of different pupils from those employed previously, if the principle of random assignment is to be satisfied.

The realities of the situation usually require that simple comparisons be used in classroom research, such as those outlined in Figure 2. Teachers should therefore resist the temptation to read backwards through the results provided in this design and to imply solutions to problems which the original design was not intended to answer. A typical design might compare the effectiveness of using two methods of instruction. Conclusions reached from such an experiment apply to the group as a whole or to other similar groups. The results may seem to suggest at the same time that girls do better with the method than boys or that it is most effective with pupils of high intelligence rather than low. However, these implications should not be stated as results or conclusions. They can be turned to advantage as statements of hypotheses or problems for further research. These implications then become the starting-points of other experiments using other similar designs to investigate these new relationships suggested by the initial study.

CHAPTER FOUR

CRITIQUE

OF

TEACHER RESEARCH

CHAPTER FOUR

A CRITIQUE OF TEACHER RESEARCH

In the final section of this report, the research studies published in Chapter Two will be examined in the light of the design and organizational problems which were discussed in the preceeding section. In particular, this section will attempt to offer some very limited evidence about the value and effectiveness of teacher research itself. Although the usual rationale for teacher research urges teachers to rely ultimately on the scientific method for evaluating instructional effectiveness, very little scientific enquiry has been spent on measuring the effectiveness of teacher research as a means of improving instruction. The teacher research projects presented here, therefore, offer some unique insights into the quality of thought which classroom research stimulates. In general, it is clear that the teachers conducting these studies have grasped the essential logic of controlled experimentation. Working under limiting circumstances of varying severity, they have constructed research which closely approximates the models for controlled comparison described in Figure 2 of the previous section. Moreover, it is apparent that these studies required considerable initiative and effort on the part of the teachers who planned them and guided them to completion.

Organization of the Studies

Under what conditions do teachers work when they do research in their classrooms and what effect do these conditions have on the research? Usually the studies are the work of one or two teachers with the cooperation of several others who participate by instructing control classes or who administer tests and provide additional data for comparison of groups. In some cases, the projects were organized with the participation of the school administration or with the staff as a group. In other instances, however, one is left with the impression that the research was largely an individual responsibility of the teacher proposing it. In these cases, teacher access to various resources such as flexibility of pupil grouping, provision of special tests, and assistance in the analysis of results is lacking.

Design of the Studies

The designs of the studies and the details of their structure are given in Table 4. For each study the following information is summarized in the table:

1. Name of the program(s) used.
2. Time over which each program was used.
3. Number of pupils assigned to control and experimental classes.
4. Grade(s) in which each program was used.
5. Use of pre- and post-tests as measures of achievement in the topics of instruction. ("Yes" or "no" indicates whether such tests were used.)
6. Use of standardized and teacher-made tests. (The kinds of tests used are categorized according to whether they are of standard or teacher construction.)
7. Use of random, matched, or equivalent groups. ("Yes" indicates that the groups designated as control and experimental were equivalent, matched or randomized by some procedure.)
8. Comparison of the gains made by the groups under programmed and teacher instruction.
9. Major conclusions of each study.

In comment on the design and organization of these studies, two points are perhaps most apparent. The first concerns the problem of establishing equivalent groups. Although most of the studies achieved equivalence of some kind between the groups before beginning the experiment, a close examination of the studies indicates that some of the matching was done "after the fact." That is to say, groups previously established on some non-random basis are compared on pre-test scores or previously collected achievement and intelligence data. Examination of these data usually leads to a decision that the groups are not significantly different in ability or post achievement. In cases where pre-existing differences are in fact detected between the groups, the researcher usually proceeds with the groups as they are and tries to compensate for observed initial differences between the groups by placing a special and usually cautious interpretation upon the results obtained in the post-tests. This difficulty in establishing equivalent groups reflects partly the teachers' restricted access to resources which would permit flexibility in grouping of pupils, and sometimes reflects a misinterpretation of randomness and matching. The effect of these restrictions on the outcomes of the studies is unknown, but may in fact be rather small.

A second comment on these studies concerns the relationship of the designs to the conclusions drawn from the results obtained. As already noted, the designs provided only for some direct and unqualified comparison between programmed and teacher instruction. In drawing conclusions from results, however, some studies interpreted the findings as though much more complex designs had been employed which permitted the drawing of conclusions regarding differential effects of the two kinds of instruction with various sub-groups of pupils.

TABLE 4. DESIGNS OF ELEVEN TEACHER-CONDUCTED RESEARCH STUDIES

Study	Programs Used	Duration of Use	No. of Pupils C	No. of Pupils E	Grade	Pre-test	Post-test	Standardized Tests	Teacher-made Tests	Equivalent, Matched or Randomized Groups	Gain	Conclusions
A	Signed Numbers	n.s.	14	14	10	Yes	Yes	IQ	Achievement; pre- and post-tests	Yes	$C \bumpeq E$	Programmed instruction served as excellent review but new topics are best introduced by teacher instruction.
	Sets	n.s.	14	14	10	Yes	Yes	IQ	Achievement; pre- and post-tests	Yes	$C > E$	
B	Ratio	11 class periods	36	36	7	Yes	Yes	IQ; Achievement pre-test	Achievement; post-tests	Yes	$C > E$	Programmed instruction was less effective than traditional teacher instruction, but experience in using it should improve its effectiveness.
C	Signed Numbers	7 class periods	n.s.	34	10	No	Yes		Achievement; post-tests	No	$C > E$	No conclusive results, though programs do vary in effectiveness. Programs are a useful method of review. This method of instruction needs special kinds of school organization to develop its possibilities.
	Ratio	8 class periods	n.s.	29	11	No	Yes		Achievement; post-tests	No	$C \bumpeq E$	

TABLE 4. CONTINUED

Study	Programs Used	Duration of Use	No. of Pupils C	No. of Pupils E	Grade	Pre-test	Post-test	Standardized Tests	Teacher-made Tests	Equivalent, Matched or Randomized Groups	Gain	Conclusions
D	Signed Numbers	n.s.	53	53	9	Yes	Yes	Reading; IQ	Achievement; pre- and post-tests	No	$C \bumpeq E$	Programmed instruction was as effective as teacher instruction and made greater provision for individual pupil differences
E	Sets	14 class periods	33	33	8	Yes	Yes	IQ; Achievement pre-test	Achievement; post-tests	Yes	$C \bumpeq E$	Programmed instruction was best suited to a role as an aid to teaching. Without supplementary teacher instruction, programs were not as effective as teacher instruction.
	Signed Numbers	14 class periods	33	33	8	Yes	Yes	IQ; Achievement pre-test	Achievement; post-tests	Yes	$C \bumpeq E$	
F	Chemistry	10 class periods	17	17	10	No	Yes	IQ	Achievement; post-tests	Yes	$C \bumpeq E$	Programmed and teacher instruction were equally effective methods.
	Chemistry	10 class periods	14	14	11	Yes	Yes	IQ	Achievement; pre- and post-tests	Yes	$C \bumpeq E$	

TABLE 4. CONTINUED

Study	Programs Used	Duration of Use	No. of Pupils C	No. of Pupils E	Grade	Pre-test	Post-test	Standardized Tests	Teacher-made Tests	Equivalent, Matched or Randomized Groups	Gain	Conclusions
G	Ratio	10 class periods	36	36	7, 8	No	Yes	IQ	Achievement; pre- and post-tests	Yes	Inconclusive	Conflicting results were received in the two grades using the programs. Programmed instruction did not exceed the effectiveness of teacher instruction and was sometimes less effective.
H	Ratio	4-7 hours	30	30	9	No	Yes		Achievement; post-tests	Yes	C > E	Programmed instruction was less effective than teacher instruction, though pupils completed the work more quickly by programmed instruction than under regular instruction.
I	Chemistry	n.s.	12	12	11-12	No	Yes	IQ	Achievement; post-tests	Yes	E > C	As a method of review, programmed instruction was superior to note review or teacher review.
J	Algebra I	November to April	133	36	9	Yes	Yes	IQ; Achievement pre- and post-tests	Pupil attitude; achievement post-tests	No	Inconclusive	The study suggested several problems in school and classroom organization which must be overcome before programmed instruction can be developed to its full potential

TABLE 4. CONTINUED

Study	Programs Used	Duration of Use	No. of Pupils C	No. of Pupils E	Grade	Pre-test	Post-test	Standardized Tests	Teacher-made Tests	Equivalent, Matched or Randomized Groups	Gain	Conclusions
K	Algebra I	October to May	18	18	9	Yes	Yes	Algebra; IQ; study habits; problem solving; critical thinking	Algebra; attitude to mathematics	Yes	$E > C$ $C \overset{\frown}{=} E$ $C \overset{\frown}{=} E$	The experimental group outperformed the control group in one of three criterion measures. For the other two tests, differences between the groups were not significant. Despite careful matching, the experimental group may have been a "stronger" class than the control group.

<u>Key to Symbols:</u>

C	Control group
E	Experimental group
$C > E$	The results obtained in the control group were significantly greater than those obtained in the experimental group.
$E > C$	The results obtained in the experimental group were significantly greater than those obtained in the control group.
$C \overset{\frown}{=} E$	The results obtained in the two groups were not significantly different.
n.s.	Not stated

<u>Descriptive Terms:</u>

Pre- and Post-tests: Tests of achievement related to a specific topic of instruction

Achievement: Tests of achievement in broad subject areas

IQ: A standardized test of intelligence

Thus, the findings of some studies relate differences of instructional effectiveness to pupil interest, attitudes, sex, intelligence and previous achievement, although the designs of the studies provide comparisons only between various forms of programmed and teacher instruction. This proclivity to wider interpretation of findings may be regarded as the reading of preconceptions about the effectiveness of instruction into the results obtained. Certainly none of the designs was well adapted to measuring differential effects between pupils and kinds of instruction. However, this weakness (if it is that) may also be interpreted as a strength. Such thinking in teacher research may go beyond the justification of formal experimental logic, but it leads also to insightful and speculative thinking by teachers about the process of instruction and the elements of which it is composed. If such thinking is not merely didactic and doctrinaire, it can be directed creatively towards hypothesis formation and later hypothesis testing through further research. Thinking that develops testable propositions is, of course, an essential aspect of fruitful research. A number of the studies clearly produced thinking of this kind. Moreover, these studies give evidence that teacher perception of the instructional process was considerably heightened and clarified. Thus, several of the studies led the participants to make incisive and pertinent comment upon learning and the conditions in classrooms which foster it.

The background of research skills possessed by the teachers conducting the studies was probably not typical of teachers in general. There are indications that most of these authors were familiar with research design and certainly with some statistical procedures before they undertook the research. Many of them also demonstrate considerable familiarity with research literature on programmed instruction. In other respects, the conditions under which these projects were conducted are similar to those which would be found in most Canadian secondary schools.

Findings and Conclusions of the Studies

The studies of programmed instruction reported in this collection set out essentially to compare programmed instruction with teacher instruction. This purpose implies that these kinds of instruction are unique and uniform entities which may be compared in somewhat the same way that two methods of growing corn might be evaluated. Such a comparison thus implies a question of fundamental importance about programmed instruction: How do programs perform in classrooms when a teacher-less environment prevails? Thus, these studies involve the issue of whether programs provide classroom instruction when teacher participation is at a minimum. Some of the studies took prior stands in regard to the issue in that programs were used only in conjunction with teacher instruction. All, however, imply that the issue of programmed versus teacher instruction was the basis of their inquiry.

The findings of the studies varied somewhat, as might be expected from the diversity of programs, schools, and conditions involved in the investigations. The results did not demonstrate the predicted superiority of programmed over teacher instruction. In approximately half of the comparisons made, the two methods were judged equally effective. In four out of 14 comparisons, however, teacher instruction was the superior of the two methods.

Two related conclusions may be drawn by examining the results of these studies and the conclusions they themselves draw. The first conclusion is that when used as an alternative to teacher instruction in the classroom, programs are no more effective than teacher instruction and may under some circumstances be less effective. A second conclusion which may be drawn is that programs are best adapted in our present school organization for use as an adjunct or aid to the usual instruction provided by a classroom teacher. Thus enrichment, acceleration, review and remedial study would be appropriate uses of this method of instruction. Both of these conclusions reflect a somewhat conservative evaluation of programs. All of these results were obtained within standard forms of school and classroom organization. Some of these studies do, however, raise the significant question of whether programmed instruction is suited to the traditional classroom at all. We are thus left with the interesting speculation as to why the programs were more effective in some circumstances than others. And there is still the further question of whether other forms of school organization are better suited than the standard classroom to meeting the requirements of programmed instruction and permitting the development of its potential.

Comparison With Other Research

Teacher research often is concerned with problems which are unique to a school or classroom. In these cases, the studies are not comparable to other research nor is there any way by which the teacher could save himself the trouble of conducting his own research by searching out previous studies conducted on the same problem. Increasingly, however, classroom research deals with matters that are of wide general interest and on which research is proceeding or has already been undertaken. Programmed instruction is an example of such an issue in education. Certainly this medium of instruction has been the subject of a great deal of comment and research. Teacher research on this same topic should therefore be related to the general framework of research in programmed instruction and should compare the findings produced by teacher research with the findings of other relevant research. By such comparison, classroom research links theory and practice and provides a genuine contribution to the scientific study of problems in instruction.

There are, of course, no final or universal conclusions to be drawn from general research on programmed instruction. Certain uniformities are, however, beginning to appear. The following statements from recent surveys of the research provide useful summaries of evidence about the effectiveness of programmed instruction.

The strategy of these studies has been to pit the classroom teacher against "the teacher in the program." Using the standard of the job being done in the schools, these studies show the teacher in the programmed textbook, by and large, to be as effective as, and more efficient than, the classroom teacher: Students learn as much in less time from programmed instruction.[1]

Do students learn from programmed instruction? The research leaves us in no doubt of this. They do indeed learn. They learn from linear programs, from branching programs built on the Skinnerian model, from scrambled textbooks of the Crowder type, from Pressey review tests with immediate knowledge of results, from programs in machines or programs in texts. Many kinds of students learn — college, high school, secondary, preschool, adult, professional, skilled labour, clerical employees, military, deaf, retarded, imprisoned — every kind of student that programs have been tried on. Using programs, these students are able to learn mathematics and science at different levels, foreign languages, English language correctness, the details of the U.S. Constitution, spelling, electronics, computer science, psychology, statistics, business skill, reading skills, instrument jying rules, and many other subjects.[2]

The findings reported in these summaries are not substantially different from those reported in the teacher research studies. The conclusions drawn from the findings do place different emphases and interpretations on the findings. The teacher researchers tend to conclude that the programs are ineffective unless they surpass the performance which would be obtained from teacher instruction. Professional researchers, on the other hand, seem not to be discouraged by findings which indicate no significant differences between the two methods of instruction. No doubt this attitude is developed by participation in the many other research studies in education which yield the same rather ambiguous outcomes. In any case, the professional researcher is likely to look for explanations of why such results are obtained and to accept the evidence of equal effectiveness between the methods as an indication that other uses or applications must be found for programmed instruction.

> The numerous experiments on programmed instruction . . . may, indeed, be proving that no significant difference exists, but the suspicion arises that in many cases the programs are too short, the samples too small, the measuring instruments too dull, to pick up differences if they exist.[3]

Thus, one of the major differences between the outcomes of teacher research and general academic research may be that teachers will tend to interpret the findings of their research in terms of the realities and limitations of present systems of school and classroom organization while the academic researcher makes interpretations according to freer terms of reference. If this situation is in fact the case, then both teacher research and academic research could profit by sharing their perceptions of the realities of the situation as they see them. To bring this sharing about, teacher research should become more coordinated, sequential and directed than is now often the case, with a conscious effort made to relate the findings and implications of academic research to the classroom. On the other hand, academic research should attempt to make concrete recommendations for change or improvement which could then be acted upon and evaluated in schools and classrooms.

Problems Suggested For Further Study

The results of these studies of programmed instruction and the comparison of their results to similar findings in academic research suggest some important problems for further investigation. First, we must ask whether programmed instruction has performed according to expectations. Secondly, we must ask whether the expectations held for programmed instruction were appropriate to its potential. In regard to the performance of the programs, the teacher research studies imply a somewhat lower level of performance than was expected. This failure to perform may be due to inadequate research procedures. Most of the programs were short; conditions varied from class to class; control and design were imperfect, though stronger than for most school evaluations; tests and statistical procedures were limited. However, some congruence among the findings in the teacher projects, and between these findings and those of academic research, gives confidence that the methods of teacher research are by no means unreliable. Other sources of explanation must therefore be sought.

An indication of a fruitful avenue of exploration is suggested by several of the teacher projects where attention was drawn to the restriction which the usual classroom organization places on programmed instruction. Thus, several studies mention classroom difficulties in the pacing of students, the provision of additional work for rapid-workers, and in establishing coherent learning in the class as a group. None of these difficulties in using programmed instruction can be solved without direct teacher intervention in classroom instruction, or without the development of new patterns of classroom organization appropriate to this medium of instruction. Thus, any either-or research procedures which view teacher and programmed instruction as two mutually exclusive activities will inevitably place unnecessary limitations upon the effectiveness of the programs themselves.

The justification for comparison of "teachers" versus "programs" may now have passed. Indeed, unless some form of teacher-less, program-centered educational institution was envisaged, the question of *either* programs *or* teachers was never a practical issue, however useful a concept it may have been theoretically. Instead, research should now be directed to discovering what combinations of teacher *and* programmed instruction yield the most effective results. Study is also needed to consider what kinds of school and classroom organization are best suited to these combinations of teacher and programmed instruction. Such studies have been called implementation studies, and while they differ from procedures of purely experimental research, they can be equally important.

1 Lassar G. Gotkin and Leo S. Goldstein. "Programmed Instruction in the Schools," *AV Communication Review*, Volume II, No. 6, November 1963, p. 277.
2 Wilbur Schramm. *The Research on Programmed Instruction.* Washington: U.S. Office of Education, 1964, pp. 3-4.

3 Schramm, p. 2.

From the point of view of school usage, it is also important to investigate variables outside the program: not only the teacher's role, but the degree of administrative support, modifications of classroom structure, and procedures for coping with differences in individual rates.

The results of such studies will not finally define the parameters of programmed instruction, but they will contribute to our scanty knowledge about the impact of programmed instruction on the classroom.[4]

Some Implications and Recommendations

A number of implications and recommendations appropriate to the foregoing discussion arise from the research reported in this publication.

Implications

These deal with questions of the feasibility, and desirability of teacher research.

1. Examination of these teacher research projects will help to justify the statement frequently made that teachers *can* do research in their classrooms. The design and organization of the studies indicated considerable ingenuity by teachers in overcoming problems. Clearly, too, these experiments show an awareness and appreciation of the requirements of controlled experimentation.

2. Research in the classroom demands considerable effort over and above the normal duties and responsibilities of teaching. Research cannot be expected, therefore, to develop fortuitously in the classroom. A number of favourable conditions are necessary to establish quality projects. Although such projects may be carried out by individuals or small groups, resources which extend beyond those normally at the command of a teacher are required. Those who value teacher research, including teachers' organizations, school administrators, and academic researchers, must cooperate to establish those conditions most favourable to teacher research.

3. Teacher-decision-making and thinking based upon classroom research are likely to be cast in a frame of reference which is determined by existing organizations of education and instruction. Many contemporary innovations in curriculum and methodology imply new structures and organization for schools and classrooms. Research concerning such new proposals should therefore provide scope for evaluation in various school and classroom conditions. Teacher and academic research should be related by systematic comparison of their findings. In this way, theoretical and practical considerations are combined in teacher research, and its value correspondingly enhanced.

Recommendations

1. Those concerned for teacher research should clarify its purposes and expected outcomes. The feasibility of engaging teachers in extensive research activity has been clearly demonstrated in the projects collected in this publication. The goals of such activity and an evaluation of means which achieve them should be considered. In the past, teacher research in Canada has perhaps been more advocated than practised. It may be time now, however, to work on problems in coordinating rather than merely promoting teacher research. How much teacher research is needed and practical? How many times should research be done on the same issue or topic? Is research the most suitable way to induce change? What relationship does teacher research bear to academic research? These and other questions of this nature should be considered now that the teacher research movement in Canada seems to be moving from a state of the possible to that of the actual.

2. Although research is often carried out by individual teachers or by small groups, the teacher researcher requires access to many resources and skills. Systematic means for providing these to teachers should be developed. Teachers' organizations are a likely source of such assistance; school administrators and public authorities are another. The response of teachers to the offer of experimental programs, and the availability of other forms of assistance, indicates how virtually any specific provisions can stimulate teacher research activity.

3. Research is inevitably research on *something*. Teachers have often been exhorted to do research. Too often these pleas are made without relating the activity to specific topics. The success of the teacher research reported here may be due in large part to the involvement of research with a particular issue — programmed instruction. Research might effectively be combined with workshops or in-service training dealing with specific problems such as team teaching, televised instruction, new curriculum, etc.

4. In training teachers in the rationale and procedures of research, attention should be given first to the provision of an adequate background from related research and academic study. That is, other relevant research should be brought to bear on the problem in which the teachers are to do their own research. The teacher researcher should then be given training in experimental design and organization of research. Finally, training should be provided in hypothesis testing and statistical methodology. In the past, this order of training has often been reversed with the result that some teacher research has suffered from an over-emphasis upon technical skills that were related neither to real problems of instruction nor to practical research designs. The order of training which emphasizes technique alone deprives teacher research of its connection with other research study in education. More important, such single emphasis obscures one of the most important goals set for teacher research — the improvement of instruction. Teacher research requires its practitioners to solve concurrently problems of meaning, design and technique. Emphasis upon one of these elements at the expense of others can only limit the value and quality of research which teachers conduct.

4 Gotkin and Goldstein, p. 281.